PERCEPTUAL EFFECTS IN PHYSICALLY BASED ANIMATION WITH RIGID AND DEFORMABLE OBJECTS

A Dissertation

by

DONGHUI HAN

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	John Keyser
Committee Members,	Scott Schaefer
	Jinxiang Chai
	Ergun Akleman
Head of Department,	Dilma Da Silva

May 2017

Major Subject: Computer Science

ABSTRACT

We perform four psychophysical studies to investigate the perceptual effect of factors in the rendering and simulation stages of physically based animation production. Our study provides helpful insights in how to improve visual plausibility or reduce computational cost, which may allow artists to adjust their designs to enhance or minimize the perceived deformation in a model, or to choose a more efficient dynamics model and simpler mesh used in simulation without harming the visual plausibility.

In our first study, we find that appearance can potentially influence people's sensitivity to differences of deformation as well as subjective rating of softness. Further analysis shows that, in simple scenarios, the effect of low-level visual details in appearance can be dominant, even if high-level information delivered by appearance has the opposite implication. Another experiment shows that as the number of objects in a scenario increases, objects are perceived to be stiffer.

In the second study, we quantitatively measure how different low-level visual details can influence people's perceived stiffness of a deformable sphere under physically based simulation. We find that checkerboard pattern with certain combinations of spatial frequency and contrast can reduce the perceived stiffness. Our study further shows that adding a high-contrast checkerboard background can reduce such effect.

In our third study, we discover that the resolution of a mesh used in the simulation of deformable objects can be reduced to a certain level without being noticed. For complex deformation, it is easier for people to recognize such reduction.

Lastly, We verify two hypotheses which are assumed to be true only intuitively in many rigid body simulations in our third study. I: In large scale rigid body simulation, viewers may not be able to perceive distortion incurred by an approximated simulation method. II:

Fixing objects under a pile of objects does not affect the visual plausibility. Our analysis of results supports the truthfulness of the hypotheses under certain simulation environments, but discovers four factors which may affect the authenticity of these hypotheses: number of collisions simulated simultaneously, homogeneity of colliding object pairs, distance from scene under simulation to camera position, and simulation method used.

ACKNOWLEDGMENTS

First of all, I would like to express my gratitude to my graduate advisor, Dr. John Keyser, for being a constant source of encouragement and insight during my research. I'm indebted to him for his endless support, motivation and patience with my research.

I would also like to thank the other members of my committee, Scott Schaefer, Jinxiang Chai, Ergun Akleman for all the time and effort they have put in. I appreciate them for the comments and suggestions they have had in my work. I am grateful for taking their courses which inspired my interests in computer graphics.

I want to thank Sergey Ozerov, Bryan Jones and Benjamin for sharing their 3D models. Without their models as a good starting point, it would take me a much longer time in modeling of all the shapes.

I also want to thank the other students that I have the opportunity to work with, Shuwei Hsu, Ruoguan Huang, Songgang Xu, Billy Clack, Daniel Miller, Josiah Manson, Lei He, Jason Smith, Hang Li and Zhipei yan. I have had a memorable time in the office with you.

Finally, I would like to express my deepest thanks to my wife and my parents for their continuous support and motivation during my studies.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a dissertation committee consisting of Professor John Keyser, Scott Schaefer and Jinxiang Chai of the Department of Computer Science & Engineering and Professor Ergun Akleman of the Department of Visualization.

Part of the stimuli animation used in the study in chapter 5 was generated by Shuwei Hsu.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported in part by NSF Grant IIS-0917286 and by Award Number KUS-CI-016-04, made by King Abdullah University of Science and Technology (KAUST).

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Motivation

Modern computer animation generated using 3D computer graphics technologies is much more realistic than decades ago. We see such change in the sloth's lifelike hair in the movie Zootopia, an AT-AT walking in a vast snowfield in the game Star Wars Battle-Front, the thrilling street shooting in Robot Recall from Oculus Rift VR and many other scenarios from the entertainment to the medical and the educational applications. It is not impossible nowadays to create a virtual scenario which people believe is real. These realistic animations generated by computers are very good approximations of their real world counterpart or a nonexistent scenario which satisfies all physical laws as in real life.

This owes much to the advancement of 3D computer graphics technologies. On one hand, the graphics hardware endows much more computing power than before. The faster hardware makes it possible to use more computation-intensive models in graphics applications. On the other hand, the graphics research community is developing algorithms and models that are more efficient, more accurate, and more controllable. Physically based dynamics models and photo-realistic rendering algorithms are among these techniques that have undergone a great improvement.

Many physically based models in computer graphics and mechanical engineering share the same physical theory. For example, Navier-Stokes equations are solved in the fluid dynamics simulations of computer graphics. They are also used to model viscous fluid in continuum mechanics. This is also true for photo-realistic rendering, articulated rigid body dynamics, deformable object dynamics, etc. In mechanical engineering, physical accuracy in the modeling and the solution is usually very important. For instance, any error above a certain level in the simulation of a bridge to be built may cause a catastrophic result. Accurate simulation in mechanical engineering often requires a huge amount of computation, meaning a server cluster or a super computer must be used. Thus, real-time computation is impossible so far in these cases. Because of this, achieving the same level of physical accuracy is often impractical in computer graphics, though researchers still strive to improve accuracy. More accurate physical animation is intuitively more realistic. For example, Finite Element Method (FEM) has been recently implemented in authoring software like Houdini for the dynamic simulation of a deformable body. This computation-intensive method further narrows the difference between the accuracy level of a deformable body animation and that of the deformation modeling in mechanical engineering. In return, the improved physical accuracy may lead to more realistic animation.

However, is a physically exact solution the ideal goal we should aim for? Researchers have noticed that extreme accuracy is unnecessary mainly for two reasons. First, there is always a certain level of discrepancy between a real world scenario and our approximation to it. Even the most accurate mechanical equipment has a level of accuracy, which means it can only avoid any discrepancy larger than the accuracy threshold. In this sense, physically based models are simply better approximations comparing to other heuristic models. They reduce the discrepancy between an approximation and its ideal ground truth by using more computing resources. Accurate models give less discrepancy at the cost of more resources. To achieve less discrepancy than that given by a method used in mechanical engineering, we may consume more resources correspondingly. However, we usually have limited resources in many computer graphics applications. Essentially, the problem is to balance the resource consumption and physical accuracy. A better understanding of a proper physical accuracy level needed is critical for any balancing scheme to be designed.

Second, subjective satisfaction, rather than physical accuracy, may be a more meaningful goal to achieve in computer graphics. Computer graphics applications, especially those for entertainment, are designed for the human viewer. People are better at perceiving animation at the conceptual level rather than judging the physical accuracy. This is mainly due to the limitation of human organs like eyes and ears. For example, people may be able to tell when a pool ball deforms like a tennis ball when it is hit. But it is very difficult for people to tell whether the separating angle of two colliding pool balls is exact or not.

Researchers have taken advantage of this observation while designing new models. In the game Star Wars The Force Unleashed, Parker and O'Brien enhanced the apparent geometric detail of the simulated objects by embedding higher-resolution, textured polygonal surfaces in a coarse tetrahedral finite element mesh [47]. This trick has been used a lot in graphics to enhance visual details. Skinning is the standard method to model human body deformation in character animations. It is not based on the physical theory of deformation, but it is efficient and gives reasonable results. Zhu and et al. considered an anatomical model of the human arm recently [66]. But, it was still far from the full musculoskeletal model of the deformable human body.

Take rigid body dynamics simulation as another example. Intuitively, it is often very difficult for people to tell whether the motion of a moving object is physically accurate or not. People may accept that a simulation result is accurate if the distortion of the object motion is small. In large scale physically based simulation, people's attention is also distracted by multiple objects. Thus it may be more difficult for viewers to tell if distortion exists. If these two hypotheses, that people don't notice small deviations and that with increasing numbers of objects people accept larger deviations, are true then a physically inaccurate simulation can still be visually plausible. To achieve higher efficiency or more controllability, many algorithms assume intuitive hypotheses like these to be true in specific scenarios. However, there is only a small amount of prior work either verifying the authenticity of such hypotheses or providing a metric to measure visual plausibility objectively.

How much physical accuracy is needed to achieve subjective satisfaction? More specifically, how does the discrepancy between an animation and its real world counterpart influence people's subjective perception of the animation? O'Sullivan was one of the first researchers to answer this question for animation with rigid body dynamics [46]. They investigated subjective tolerance on several kinds of distortion in physically based rigid body simulations. Even though their observation was limited to rigid body dynamics and the discrepancy in their work was only from artificial distortion, they provided meaningful insight in this direction.

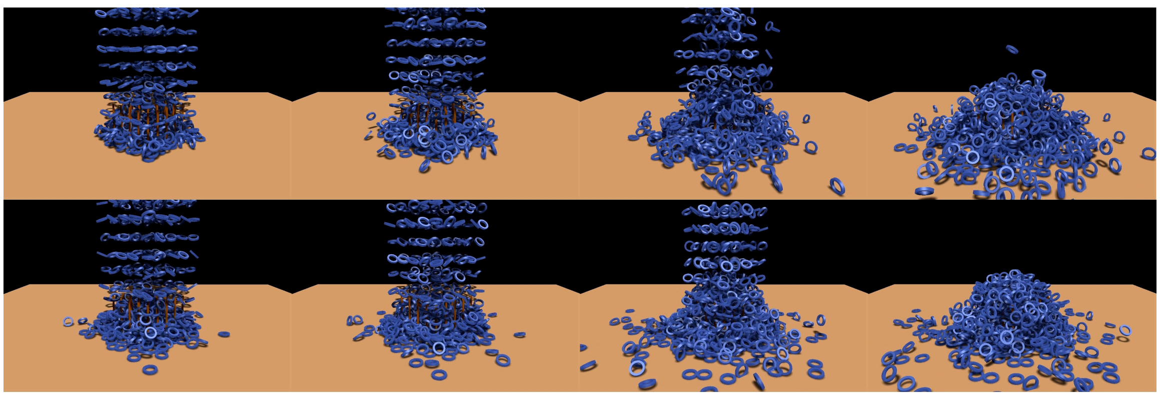


Figure 1.1: Screen shots from two stimuli used in our study. Screen shots in upper row are from a simulation using classic impulse based rigid body dynamics model, while screen shots in lower row are from a simulation with statistical simulation method [27].

The work summarized in this dissertation is to go a step further and to answer the above question in more details. Both the discrepancy and people's subjective perception need to be investigated carefully before a formal study can be performed. The discrepancy between an animation and its real world counterpart come from different sources. Almost every stage in animation production can incur discrepancy. The commonly seen sources of discrepancy include dynamics models, dynamics model parameters, simulation meshes, rendering meshes, rendering textures, complexity of scenarios, etc.

The discrepancy from some sources are difficult to control, for example, the discrepancy caused by dynamics models. Different dynamics models have different levels of accuracy. Dynamics models are usually categorical and cannot be easily replaced by a different one once selected. Figure 1.1 shows an example of a dynamic scenario simulated using two different dynamics models. They have different accuracy levels thus different discrepancy.

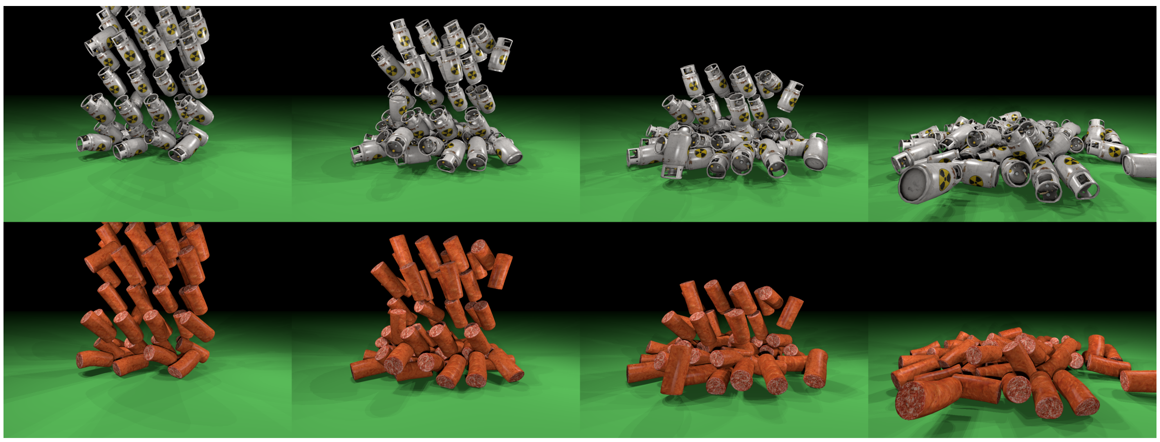


Figure 1.2: Frames taken from two animations used in our study. Both animations are rendered using the same physically based simulation of 64 deformable tetrahedral models in the shape of cylinder. Two appearance(gas tank and sausage) are used for rendering.

The discrepancy from sources like rendering meshes and rendering textures are easier to control and tune. Figure 1.2 shows an example of a deformable body dynamics simulation rendered using different rendering meshes and rendering textures. Every object in the scenario is a deformable object. The real stiffness and other material properties of the model under simulation may be completely different from a gas tank or a sausage. Therefore both animations have discrepancy caused by the appearance. Also, their discrepancy can be at different levels. But, in this case, it is easier for us to alter the appearance to control the discrepancy.

Factors from sources other than the major animation production stages can also influence the discrepancy. For example, the complexity of scenarios, more specifically the number of objects in a scene or the complexity of the background, can increase or reduce the overall discrepancy.

People's subjective perception of animation can be evaluated from different perspectives. The visual plausibility is just an overall and vague perspective of the subjective perception. However, it serves as a good and comprehensive metric in an elementary study. The perceived softness of an object, for example, is another aspect of people's perception of physically based animations. Metrics of subjective perception like this are more specific and allow us to perform numerical analysis. In our study, we consider both general and specific metrics.

Our concrete objectives are to answer the following five questions. *What factors can have significant influence on people's perception of physically based animations? What aspects of people's perception can these factors influence? How do different factors influence these aspects of people's perception? What other factors can influence the effect of these major factors? How can we make use of these effects?* Obviously, these general questions can be asked in many areas in computer graphics. In our work, we only focus on perception of animation with physically based rigid body dynamics and deformable body dynamics. We consider discrepancy from both rendering and dynamics simulation. We also study other factors which may influence the effect of the main factors. We design multiple scenarios to allow the result to be more general. We use the visual information as the sole feedback. So the result is not influenced by other feedbacks like haptics or audio.

Once a good understanding of how the discrepancy can influence people's perception is available, one can speculate how much discrepancy is acceptable to maintain subjective satisfaction. The knowledge can be useful even in applications with enough resources for dynamics computation and rendering. Rather than simply using the most physically accurate implementation, we can determine how much effort we need to achieve subjective satisfaction and save the remaining resources for other purposes. The knowledge of such influence can be used by artists in a more proactive way in authoring. For example, Held and et al. presented a probabilistic model of how viewers might use defocus blur in conjunction with other pictorial cues to estimate the absolute distances to objects in a scene [25].

1.2 Thesis Statement

My thesis statement is:

The factors in the simulation and rendering stages of the animation production pipeline can significantly influence the perception of physically based animation of deformable and rigid objects. These factors can be controlled to enhance or reduce such influence to benefit the animation production.

Although many factors can influence the discrepancy of an animation, we mainly study four factors: the appearance, the mesh resolution of a deformable model, the dynamics model in rigid body simulation and the complexity of scenarios. Each factor may influence different aspects of the perception. We aim to prove that the appearance of an object can significantly influence the perceived stiffness and the perceived difference of deformation. More specifically, we plan to investigate what information in the appearance determines the effect. To prove that such effect can be controlled to be beneficial, an effort should also be made to fit a numerical model for such effect. Next, we want to show that the reduction of the resolution of the simulation mesh can significantly influence the perceived difference of deformation. By trying to fit a numerical model, we want to show that the reduction is significantly noticeable only when it is below a threshold. We further investigate what aspect of deformation can influence the threshold. Similarly, we plan to show that a viewer may not be able to perceive the discrepancy incurred by approximated rigid body dynamics models in a large scale rigid body simulation or a simulation with piling.

In the following chapters, we will clarify the context and give the definition of each factor mentioned above as well as how we measure the different aspects of the perception. From there on, we prove each specific statement by performing user studies and statistical analysis. So, the result is usually reported as whether a factor has a statistically significant effect in an experiment. The detailed analysis on these effects and how they can be beneficial will be given in each study separately.

In each study, we first quantify the determinate factor and choose correspondingly the quantitative representation of the subjective perception. A set of experiments are then designed and performed to collect subjective feedback. By statistically analyzing the data, we can tell whether a factor can significantly influence the quantitative representation of the subjective perception. It is also possible to know what the influence is like from the data. We study rigid body dynamics and deformable object dynamics separately. We also study the rendering and the simulation factors separately. Thus, we can have clear and fundamental conclusions for each factor that we are interested in. Once they are studied independently, we can draw an overall conclusion about the thesis statement.

1.3 Research Overview

To prove the thesis, we've performed four studies which we denote as ***Study I~IV*** from here on. Much of this work has appeared in recent publications [21, 22, 23].

We first study the discrepancy from appearance [22] of deformable objects in ***Study I***. We consider deformable object dynamics because the appearance is more likely to have influence on deformable objects than on rigid bodies. We use the perceived difference in deformation and the perceived stiffness as more concrete metrics of the subjective perception. As a primary study, we try to investigate the effect of appearance on people's perception of physically based simulations of deformation. We consider both high-level conceptual information and low-level visual details in appearance. We ignore the source of low-level visual details and only qualitatively compare appearance with more or fewer low-level visual details. In two psychophysics experiments, we analyze to see which between high-level conceptual information and low-level visual details is more dominant in effect. We apply a number of different appearance and rendering parameters to these objects, and then use two user studies to measure whether appearance used for an object can have a statistically significant effect on the perception of its deformation. In another study, we adjust the number of objects simulated and investigate how this can influence the effect of appearance. We find that appearance does influence peoples sensitivity to difference of deformation as well as the subjective rating of stiffness. Further analysis shows that, in simple scenarios, the effect of low-level visual details in appearance is dominant, even if high-level information delivered by appearance has an opposite effect. The third study shows that as the number of objects in a scenario increases, objects look stiffer although the Young's Modulus is not changed. The effect of low-level visual details is weaker. High-level conceptual information seems to take more effect.

To better understand the influence of discrepancy from appearance, we perform ***Study II*** [23] to quantitatively measure how different low-level visual details can influence people's perceived stiffness of a deformable sphere under physically based simulation. We use a checkerboard texture to render the simulation of a free falling sphere that collides with the ground and bounces up. We vary the spatial frequency and the contrast of the checkerboard pattern according to results seen in a previous study on the Spatial-Temporal Contrast Sensitivity Function (CSF). We also add a high contrast checkerboard background to study how complex backgrounds can influence the effect of low-level details of foreground objects. We prove the significant effect of spatial frequency and the contrast on people's rating of stiffness of a sphere textured with a checkerboard pattern. Furthermore, we find certain combinations of the spatial frequency and the contrast can reduce the perceived stiffness. The effect is larger for stiffer spheres. We discover that backgrounds with high contrast checkerboard textures can reduce the effect of foreground low-level visual cues on spheres. Background textures can potentially interfere with people's perception of deforming textures that have similar spatial frequency and contrast. The result can be used to create a metric for artists in designing textures to enhance or reduce the stiffness perceived by a viewer.

In ***Study III***, we investigate the more interesting factors in the simulation stage of animation with deformable objects. Because the simulation of deformable objects is more complex and time-consuming, researchers have proposed different methods to improve the efficiency. The advantages and weakness of these methods have been discussed as well. We choose to study the resolution of the simulation mesh as a factor which influences the discrepancy of the simulated animation directly. The mesh resolution can be controlled by the artists easily and directly without requiring the manipulation of the dynamics engines. And it is directly related to the computational cost of the simulation. So it is practical to adjust this factor to achieve the expected balance between visual satisfaction and simulation efficiency. We aim to provide measured data of subjective feedback to give more insight into the effect of adjusting the mesh resolution. We design two experiments using eleven different scenarios and eight different shapes. Each shape has six simulation meshes at different mesh resolution levels. In each scenario, we render using the same shape but simulate using the corresponding six simulation meshes to create similar animations. Viewers are asked to observe the difference of the final animations. Our data shows that viewers only notice the difference caused by the reduction of the mesh resolution with a high probability when the mesh resolution is reduced to a certain level. Furthermore, the reduction level at which people can notice the change is influenced by the complexity of the deformation in the simulation using the simulation mesh of the highest resolution. Our observation proves the effectiveness of the reduction of the model resolution in improving the simulation efficiency without harming the visual satisfaction. However, our analysis also suggests analyzing the complexity of the deformation to determine a proper level of reduction of the mesh resolution.

Lastly, in ***Study IV*** [21], we choose large scale rigid body simulation as the scenario. Rigid body dynamics models are simpler than deformable models. The resulting scenario is also simpler. We choose dynamics models and the scene complexity as sources of the discrepancy. Visual plausibility is used as the representation of subjective perception. The visual plausibility of scenarios simulated is measured using the subjective rating from viewers. To further investigate what factors may affect the results, we carry out another group of experiments to measure the effect of four other factors: the number of collisions under simulation, the homogeneity of colliding object pairs, the distance from the scene to the camera, and different simulation methods. In this study, we had the following conclusions. In the large scale rigid body simulation, depending on the dynamics model used, viewers may or may not be able to perceive the discrepancy incurred by an approximated simulation method. Assuming objects in or under a pile of objects to be fixed without transformation does not affect the visual plausibility of simulation. It is obvious that number of objects does have an influence on people's evaluation. It is more difficult for people to tell the difference among stimuli using different physical models when the number of objects under simulation is very high. We also explore whether eye-tracking can provide an objective metric of visual plausibility. Unfortunately, we find no evidence that this is the case.

1.4 Section Overview

The following chapters in this dissertation are organized as following. Section 2 reviews the background and the related work. This mainly covers basics of various dynamics models used in our studies. It also surveys related work in perception in computer graphics, especially in the perception of animation with dynamics. Section 3 discusses our studies of the discrepancy caused by rendering factors of deformable object simulation. The study on overall appearance is discussed first. Then we focus on low-level visual details. Section 4 summarizes our investigation of factors in the simulation of deformable objects. Section 5 discusses our studies of the discrepancy caused by factors in the simulation of large scale of rigid bodies. Section 6 concludes this dissertation with a summary, discussion of limitations, and notes for directions of future work.

2. BACKGROUND

2.1 Key Factors in The Production Pipeline of Physically based Animation

The entire production pipeline of a computer animation includes stages like story boarding, concept design, layout, voice recording, modeling, texturing, rigging, animation, special effects, lighting, rendering and other post production stages. Since we are only interested in physically based animation, we focus on the related stages like modeling, texturing, special effects and rendering. The physically based simulation happens in the special effects stage. From another point of view, these stages determine the appearance of an object or a character in an animation and their dynamical motion over time. We start our work by investigating factors like the appearance and the dynamics models in these stages. We first review basic concepts and related works about these factors.

2.1.1 Rigid Body Dynamics Models

Rigid body dynamics has been well studied. For an introduction of the elementary concepts about rigid body dynamics, we refer the recent textbook [14]. For an overview of concepts and technologies in the area of rigid body simulation, the state-of-the-art report [4] is a good material. Here we briefly review some work on large scale rigid body simulation which we will use in the study.

Multi-body interaction is commonly seen in real world scenario (e.g. billiards, explosion, building collapsing, etc.). Large scale rigid body simulation is of interests when physically based modeling of such phenomenon is necessary. Obviously, the more objects being simulated at the same time, the more computation needs to be done meanwhile.

Large scale rigid body simulation is different from the simulation of a single object in terms of the amount of objects moving at the same time. What attracts viewer's attention in these two scenarios can be different too. The movement of objects in the global level in large scale rigid body simulation, which is not available in the scenario with only a few objects, can be an important phenomenon that determines the subjective plausibility of animation. Inspired by such observable phenomenon in large scale rigid body simulation, Hsu and Keyser proposed statistical model [27] to approximate the accurate collision detection/resolution method. This approximation can reduce the simulation time of one time step to as low as 1% of a full simulation step. Similarly, they approximated the piling phenomenon of large scale rigid body simulation by "freezing" the simulation of nearly static objects that satisfy certain criteria [28] or adding constraints that help maintaining the pile shape [29].

In these work, approximation or simplification of fully physical models are used to save computation. The effectiveness of the trade off between physical accuracy and computation speed is based on the assumption that the reduction in physical accuracy does not incur much reduction in visual plausibility.

2.1.2 Deformable Body Dynamics Models

Nealen and et al. summarized the improvement of physically based deformable models in computer graphics in recent years [44]. An earlier survey by Gibson and Mirtich can be found in [19].

Among the several popular deformation models used in computer graphics, we choose the Lagrangian mesh based methods, which are well-developed and widely used, to simulate most of the deformable objects in our stimuli. FEM based methods constitute a big part of this class of methods. They are based on a detailed discretization of traditional mechanics theory. Also, people have optimized the performance of this class of methods in different aspects. Therefore, they are more stable, reliable and accurate. Thus, they can provide realistic simulations.

In continuum elasticity theory, the undeformed shape of an object is usually a continuous connected domain $M \subset \mathbb{R}^3$. The coordinates of an undeformed point $\mathbf{m} \in M$ is called the *material coordinates* of that point. In the deformed configuration, the same point locates at its *world coordinates* $\mathbf{x}(\mathbf{m})$. The displacement from *material coordinates* to *world coordinates* $\mathbf{u}(\mathbf{m}) = \mathbf{x}(\mathbf{m}) - \mathbf{m}$ is a vector field defined on M. The displacement gradient is a simple measurement of the variation of the displacement field.

$$\nabla \mathbf{u} = \begin{bmatrix} u_{,x} & u_{,y} & u_{,z} \\ v_{,x} & v_{,y} & v_{,z} \\ w_{,x} & w_{,y} & w_{,z} \end{bmatrix} \tag{2.1}$$

Two commonly used strain tensors in mechanics and computer graphics are the Green-Saint-Venant strain tensor ε_G and Cauchy's strain tensor ε_C.

$$\varepsilon_G = \frac{1}{2}(\nabla \mathbf{u} + [\nabla \mathbf{u}]^T + [\nabla \mathbf{u}]^T \nabla \mathbf{u}) \tag{2.2}$$

$$\varepsilon_C = \frac{1}{2}(\nabla \mathbf{u} + [\nabla \mathbf{u}]^T) \tag{2.3}$$

To integrate the status of the system to the next time step, we need to know the internal force given the strain tensor in the entire domain M. The stress tensor is a fundamental force descriptor from which we can compute force density and traction. Two commonly used stress tensors are Cauchy's stress tensor σ and the Piola-Kirchhoff stress tensor P. A constitutive law builds the connection between the strain tensors and the stress tensors. Different derivations can be used for different materials. For elastic materials, Hooke's law can be used:

$$\sigma = C \cdot \varepsilon \tag{2.4}$$

For hyper-elastic material, the derivative of deformation energy density with respect to the displacement gradient can be used. For isotropic materials, this connection is influenced by two independent values, Young's Modulus E and Poisson's Ratio ν. Between these two, Young's Modulus is a measure of material stiffness. The higher this value, the higher stress is caused for a specific strain. In our study, we adjust this parameter to achieve different stiffness.

2.1.3 Texturing and Rendering

In the modeling and texturing stage, artists determine the exterior appearance, like shape, color and texture, of an object or a character. Also the color information is mapped from the 3D object onto a 2D texture plane. But this information is used later in the rendering stage. In our study, we treat the appearance as a factor in the rendering stage.

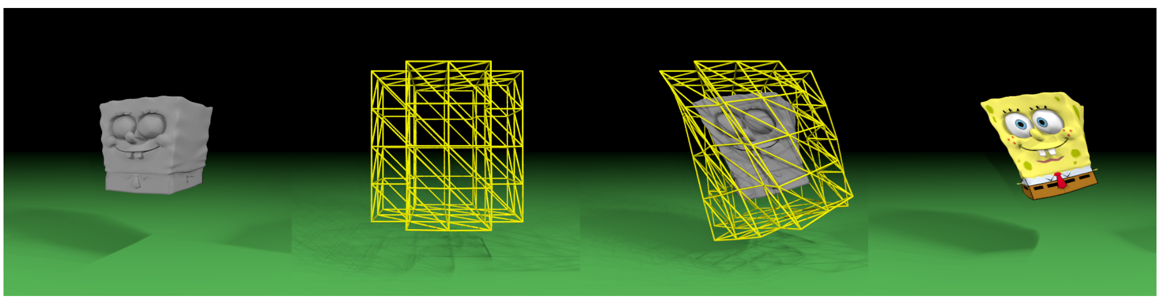


Figure 2.1: Example of the input and output of the special effect(simulation) stage and the rendering stage. From left to right: the rendering mesh, the simulation mesh, the simulated result with the recovered rendering mesh and the rendered object.

Figure 2.1 shows an example of the shortened pipeline from the modeling to the simulation and the rendering stages. Given a model as shown in figure 2.1 leftmost, a separate simulation mesh can be generated as the yellow mesh. The simulation stage uses the dynamics models discussed in previous sections to determine the status of the simulation mesh on the fly. The rendering model can then be recovered using methods like interpolation. Given this output of the simulation stage, the rendering stage then apply the 2D texture on the deformed rendering mesh and compute the final image on the image plane.

Apparently, the appearance and the dynamics model of an object determines different aspects of the final animation in separate stages. Thus, we study them as two independent factors.

2.2 Plausibility of Computer Generated Animation

2.2.1 Physical Plausibility

Physical plausibility was investigated in early research on developing metrics for measuring simulation plausibility. Yeh and et al. tried to propose an embracive metric which measured error between a distorted simulation and a baseline simulation [65]. Their metric was a combination of seven global or instantaneous metrics, including energy difference, penetration depth, linear/angular velocity magnitude, etc. However, it was still not a general one due to their specific error sampling method. Also, it measured only physical accuracy not perceptual accuracy. Also, simulation which had intentionally injected error for aesthetic reason might have a low rate by this metric. But it might be preferred by people. As a result, this metric might not be as practical as proposed.

2.2.2 Visual Plausibility in Computer Graphics

On the other hand, visual plausibility is measured extensively in evaluating simulation and modeling in a broad range of fields. Visual plausibility is a proper choice especially in entertainment applications like movie and game where the goodness of the product is judged by human subjects eventually. Intuitively, there could be discrepancy between physical plausibility and subjective visual plausibility due to the characteristics and limitation of human perception. Using visual plausibility directly avoids the need to consider such discrepancy in these applications.

One popular application of this metric is in measuring the quality of Level-Of-Detail (LOD) techniques. For example, McDonnell and et al. presented a perceptual evaluation of different LOD representations of humans wearing deformable clothing [38]. Larkin and O'Sullivan explored the perception of texture, silhouette and lighting artifacts on simplified character models [34].

Human posture can provide a significant amount of important information through nonverbal communication. It shows whether a person is happy, angry, aggressive, etc. It is humans instinct to recognize such nonverbal message. In turn, it is very natural and common to verify the quality of animated character motion and crowds by measuring their visual plausibility. Reitsma and Pollard measured viewer sensitivity to errors in animated human motion like jumping [53]. Pražák and O'Sullivan showed that people can perceive even very low levels of foot sliding through psychophysical experiments [51]. Vicovaro and et al. studied people's sensitivity to manipulations of overarm and underarm biological throwing animations [60]. Dynamic Time Warping is often used to edit the biological throwing motions. This work investigated the effects of such modification on visual plausibility. Hoyet and et al. investigated the effect of anomalies in physical contacts between characters and matching appropriate reactions to specific actions [26]. They found that timing errors, force mismatches and angular distortions were all easy to detect for human.

In scenario with more than one character avatars, the interaction among avatars and that between avatar and the environment require more factors to be considered while studying the visual plausibility. McDonnell tested different factors which may affect the perception of variety of crowds [39]. They found motion clones were more difficult to detect than appearance clones. They further investigated which body parts of virtual characters were most looked at in scenes containing duplicate characters [40]. McDonnell and O'Sullivan then investigated how audio and visual cues affected human in determining sex of conversing characters [41]. Ennis and et al. measured the effect of audio mismatches and visual desynchronization on visual plausibility of conversing characters [15].

Similar to physically based dynamics, it is common for physically base rendering to trade off between physical accuracy and computational efficiency. Measuring the visual plausibility helps to verify if the improvement in computational efficiency is at the cost of large reduction in subjective plausibility. McDonnell and et al. examined whether render style affects how virtual humans are perceived [37]. Jarabo and et al. proposed a metric to evaluate the visual plausibility of illumination in scenes with dynamic aggregates [31]. They also showed that factors like complexity of geometry and temporal properties of the crowd entities, the motion of the aggregate as a whole, and the presence or absence of color could all affect perceived fidelity. Harrison examined to what extent the lengths of the links in an animated articulated figure could be changed without a viewer becoming aware [24]. Vangorp and et al. performed visual perception experiments in developing a predictive model of distortion for street-level image based rendering [59].

The visual plausibility has also been studied in latest applications like virtual reality [8, 13]. Piovarči and et al. addressed the problem of mapping a real-world material to its nearest 3D printable counterpart by constructing a perceptual model for the compliance of nonlinearly elastic objects [50].

2.2.3 Visual Plausibility of Physically Based Dynamics

Even though the study of visual plausibility has not been nearly as widely explored in physically based dynamics research, visual plausibility has been adopted and integrated into dynamical simulations to achieve specific simulation effects. Chenney and Forsyth extended classic simulation models to include plausible sources of uncertainty and then used a Markov Chain Monte Carlo algorithm to sample multiple animations that satisfy constraints, which could include visual plausibility [9]. Similarly, Twigg and James exploited the speed of multi-body simulators to compute numerous example simulations in parallel [58]. It allowed viewers to modify the simulation interactively according to its visual plausibility.

Garcia and et al. verified through perceptual evaluation that stylized motion, although it might incur distortion, could improve visual plausibility of physically based simulation [17]. In another study, they evaluated the visual plausibility of their deformation simulation [18]. They further studied whether deformation helped in detecting collisions. But again, only a few objects were involved in their scenarios. Rather than measuring the distortion caused by the approximated algorithm, Reitsma and O'Sullivan examined how a realistic environment setting could affect visual plausibility of physically based simulations [52].

These work are usually constrained to a very specific application or purpose. In our research, we aim to approach the more general problem of the effect on perception of physically based animation.

2.3 Perception of Dynamics in Animation

2.3.1 Perception of Rigid Body Dynamics

O'Sullivan and et al. was the first to evaluate the visual plausibility of simulations where physical parameters had been distorted or degraded, either unavoidably or intentionally for aesthetic reasons [46]. They found some interesting biases as well as derived a set of probability functions which could be used as a metric to evaluate the visual plausibility of simulation. However, their work was limited in the specific and simple scenarios. Their discovery could not be easily generalized to other scenarios, for example, animation with multiple objects with uncontrollable discrepancy.

2.3.2 Perception of Elastic Deformation

People have studied the impact of visual information on the haptic perception of mechanical stiffness in virtual environments. Srinivasan and et al. performed experiments to study the effect of visual feedback on people's discrimination of stiffness of virtual springs [56]. Their results demonstrated a clear visual dominance over kinesthetic sense of hand position. Moody and et al. demonstrated that a single physical surface could be made to “feel" both softer and harder than it was in reality by presenting visual information [42]. Warren et al. studied the influence of kinematic patterns of object motion in visual and auditory perception of elasticity in bouncing objects [62]. These studies show that visual feedback is profitable to haptic perception.

How visual information as a sole feedback can influence people’s perception of deformation is a more fundamental problem. Due to the complexity of dynamics modeling and the hardness to numerically model the factors like visual information, there were only limited work on this topic. Argelaguet and et al. introduced the so-called Elastic Images, a pseudo-haptic feedback technique which provided perceivable and exploitable sensation of images without the need of any haptic device [2]. Elastic local deformation of images, procedural shadows and creases were used as visual cues. Pejsa studied which properties of deforming objects people perceived, and which vocabulary terms they used to describe these properties [49]. García and et al. designed a perceptual experiment to test whether the appeal of deformation simulation was improved by their method and studied local deformations effect in perception of contacts [18]. Kawabe and Nishida studied human perceived elasticity of transparent cube being simulated as a deformable object [32]. Their results suggested that human elasticity judgment of transparent cube is based on the pattern of image motion arising from contour and optical deformations. Aliaga et al. explored some factors that contribute to the perception of cloth, to determine how computation efficiency could be improved without sacrificing realism [1]. They considered the appearance and dynamics as two interplay factors in determining the perception of animated cloth. Rather than focusing on one specific scenario like cloth simulation or transparent soft body simulation, we tried to study the more general case of animation with deformation. And we aimed at more fundamental effects of appearance and dynamics by limiting their mutual interference and studying them in separate experiments.

2.3.3 Perception of Low-Level Visual Details

A good numerical model of the visual information in animation is the key to better understand the influence of visual feedback on people's perception. Comparing to visual conceptual information in the animation, which is more difficult model, a bottom-up model that uses low-level visual details is more straightforward to use in the perception study.

Low-level visual details can be defined and measured in different ways. Norman and et al. proved the importance of low-level details like specular highlights and shading in perception of 3D shape [45]. Sweet and Ware investigated the relationships between view direction, texture orientation, and surface orientation in surface shape perception [57]. Winnemoller and et al. presented a psychophysical experiment to determine the effectiveness of various low-level shape details for rigidly moving objects in an interactive, highly dynamic task [63].

Comparing to the above definition of low-level visual details for specific scenario, we need a general model that can represent visual information in more cases. Sensitivity of the human visual system to low-level details of different spatial and temporal (flickering) frequencies had been measured to generate the Contrast Sensitivity Function (CSF) [54]. CSF indicated that people are most sensitive to middle frequency gratings. $(e.g. 8 cycles/degree)$. Kelly further studied people's sensitivity to moving gratings of different spatial and temporal frequency and fitted a model of the contrast sensitivity surface on the spatial-temporal frequency domain [33]. That study was based on stabilized stimuli whose velocity was measured as retinal velocity. Daly revisited the fitted model and extended the model to image plane velocity [12]. This allowed the result to be used more conveniently in applications where an accurate eye tracker was not available. The CSF model had been proven to be useful in different research areas. An early study by Mannos and Sakrison made use of the CSF for measuring the visual fidelity of monochrome still images [36]. Yee and et al. accelerated global illumination computation in dynamic environments by taking advantage of the CSF [64].

However, there is no prior work we know of studying the sensitivity of the human visual system to low-level details when evaluating deformation using CSF of rigid textures. As will be discussed in following sections, we try to make use of the CSF to quantitatively study the effect of low-level visual details in people's perception of deformation.

3. FACTORS IN THE RENDERING OF ANIMATIONS WITH DEFORMABLE OBJECTS*

We first investigate the effect of the appearance on people's perception of simulation of deformable objects. The appearance is one the most important factors in the rendering stage of the animation production pipeline. It determines in a great extent most of the visual information that people perceive. Comparing to factors like dynamics models or lighting models, the appearance of objects is easier to control. More specifically, the influence of the appearance on perception, if there is any, can be finely tuned by carefully designing the details of textures or the details of rendering meshes. Therefore, the result of our study on the appearance can be most interesting to people. We first discuss our work on the appearance in this chapter.

3.1 Motivation

Due to the limited previous work on this topic, we approach this problem in two steps. In Study I, which is the elementary stage, we design some tentative experiments in order to qualitatively understand whether and how the appearance take effect. Based on the knowledge learned from Study I, we perform Study II, which aims at gaining quantitative analysis on the effect of the appearance. In each study, we perform several independent experiments to collect data and perform statistical analysis. We first discuss the motivation of both studies in this section and then give more details about the two studies in following sections.

---

*Part of the data reported in this chapter is reprinted with permission from "Effect of low-level visual details in perception of deformation" by D. Han and J. Keyser, 2016. *Computer Graphics Forum*, volume 35, pages 375-383, Copyright 2016 by John Wiley & Sons, Inc. Part of the data reported in this chapter is reprinted with permission from "Effect of appearance on perception of deformation" by D. Han and J. Keyser, 2015. *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 37-44. Copyright 2015 by ACM.

3.1.1 Study I: Effect of Overall Appearance of Objects

3D modeling and texturing are important stages in 3D animation production pipeline. They directly determine the appearance of an object or a character, and the information delivered by the appearance. They can potentially influence the visual plausibility of the object or character in rendered animation. In Fig. 3.1, the same deformed tetrahedral model of a torus is rendered with two different embedded rendering meshes and textures. By looking at the static images separately, one may have the impression of a more rigid rim, even if it looks plastic and damaged, and a softer tire. Such judgment is based on the ***high-level conceptual information*** delivered by the appearance of an object—*the overall appearance of the model as interpreted by a viewer's prior knowledge*. Although such high-level concepts can provide much information, we focus in this study on its effect on perception of material deformability. Just as the modeling and texturing of an object is used to give a viewer the sense that they are seeing an object from real life, it is intuitive to think that such high-level conceptual information will also influence perception of the deformability of the object. However, as we will see from our studies here, this is not an accurate assumption.

In animation, spatial-temporal information is available. Research on the ***low-level visual details***, *such as high-contrast(salient) boundary or corner point*, has proved their importance in people's visual cognition since the work of Robson [54]. Low-level visual details of various amounts and in different forms can have potentially significant influence on people's perception of deformation. We are faced with a question of whether high-level conceptual information can still have a dominant significant influence on people's perception. In other words, does the rim still look more rigid than the tire in animation?

O'sullivan and her colleagues have investigated people's sensitivity to distortion in physically based simulation[46, 18]. However, there is little study on how low-level visual

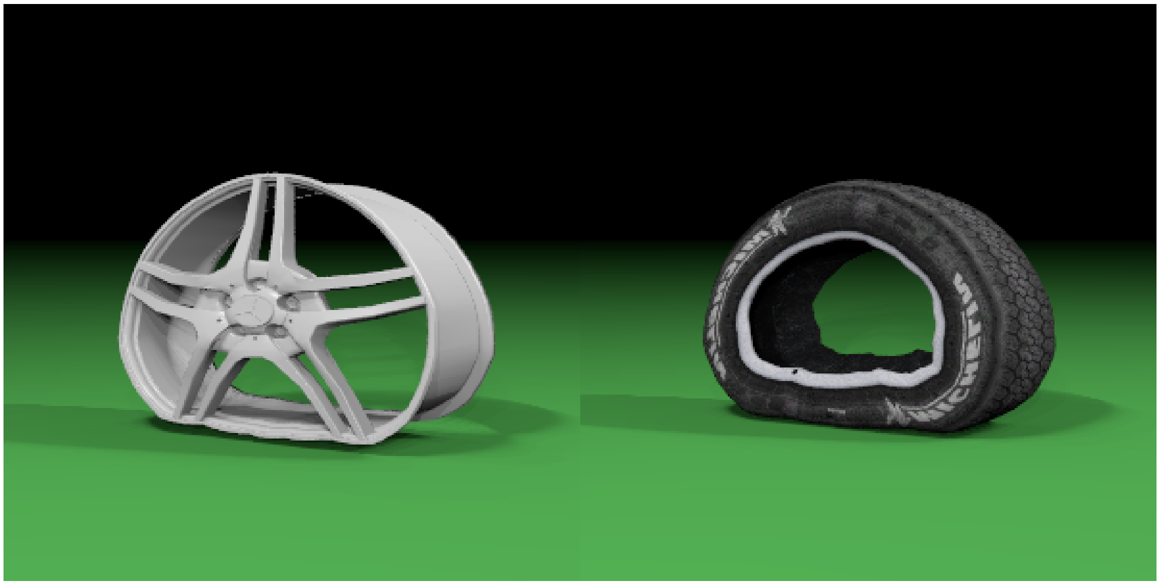


Figure 3.1: One frame of simulation of a torus model rendered as rim (left) and tire (right)

details and high-level conceptual information influence people's perception of deformation interactively. This study can potentially help artist to adjust the design of appearance to balance between faithful high-level conceptual information and specifically tailored low-level visual details if they want to enhance or reduce the deformation perceived by people.

3.1.2 Study II: Effect of Low Level Visual Details

As will be discussed in detail in following sections, our Study I shows that there is a significant effect of the appearance on people's perceived stiffness of deformable objects in simple scenarios with only a few objects and simple backgrounds. Study I further shows that low-level visual details, rather than high-level conceptual information, in a texture are more dominant in influencing people's perception in simple scenario.

In Figure 3.2 top left, for example, it is the low-level visual details like the curved edges between a black pattern and white color on the sphere that dominates the influence, rather than the high-level hint of "a deformable soccer ball". In Study II, we quantitatively

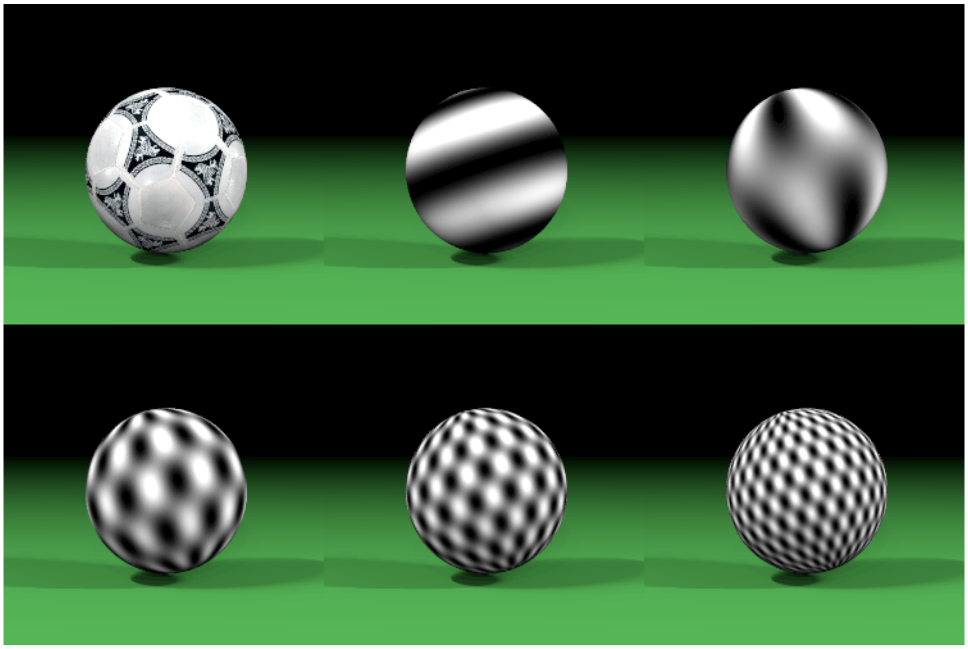


Figure 3.2: Sphere rendered using soccer texture (top left) and using textures inverse-transformed from five major components of Discrete Fourier Transformation of the soccer texture.

measure the effect of such low-level visual details.

Kelly has studied the sensitivity of the human visual system to moving rigid low-level visual details in depth [33]. The sensitivity is measured by how well people recognize the color or intensity variation in grating stimuli. The sensitivity is found to vary for grating stimuli of different contrast, spatial frequency, and moving speed. We take advantage of this result by quantifying low-level visual details in a similar manner. This method allows the possibility to combine our result with a Discrete Fourier Transformation (DFT) or a Wavelet Transformation (e.g. Harr wavelet) to provide an overall estimation of people's perceived stiffness for any texture. For example, the soccer appearance in Figure 3.2 can be decomposed into a weighted summation of the other five textures shown, each of which only contains low-level visual details of a sinusoidal pattern with specific frequency and contrast (controlled by weight). Knowing the perceived stiffness for each component, an estimation for the overall appearance can be computed.

In our primary study, we make several simplifications in order to make the quantitative study practical. Instead of using sinusoidal texture as in previous studies, we use a checkerboard texture, which can also serve as fundamental basis. DFT decomposes a checkerboard texture into a fundamental sinusoidal component, whose frequency equals the spatial frequency of the checkerboard, and other sinusoidal components with multiples of the fundamental frequency. The higher-frequency components, although only constitute a small fraction in the energy spectrum of a checkerboard texture, make the tracking task easier. This helps us to discover the effect of low-level visual details if any. Also, the result can guide future studies using sinusoidal textures. Because the fundamental sinusoidal component constitutes the most significant part in a square wave, we approximate the Contrast Sensitivity Function(CSF) of a checkerboard stimuli using the CSF of a sinusoidal stimuli whose frequency is equal to the fundamental frequency. We use the approximate CSF both in determining a proper spatial frequency-contrast sampling domain heuristically and in understanding the measured data. Although there could be discrepancy between the CSFs of these two stimuli, our result still reveals obvious pattern as discussed later. However, due to the higher-frequency components, our measurement can not be directly combined with DFT in stiffness estimation. It may be more convenient to use wavelet transformation in this case.

When measuring frequency, we use the frequency for the undeformed object, although in practice the spatial frequency of a texture changes as the object deforms. We also do not study the effects of object shape (i.e. we use only spheres) or color variations, both of which would be important in more realistic images. We quantify people's perception of deformation using their subjective rating of stiffness of a deformable object in simulation. The rating in our study has five integral levels corresponding to five reference stiffness levels. Because people base their rating of stiffness level mainly on the apparent deformation observable. The deformation model and other simulation parameters can all influence the extent of deformation and people's judgment. However, the deformation extent and the stiffness are negative correlated in our study. Also, the relative difference in people's rating, rather than the absolute rating level, is more critical in the statistical analysis. In our study and discussion, we use "rating of stiffness" for consistency. As our result shows, even with variation in people's subjective ratings, statistically significant patterns can be observed.

3.2 Study I Experiments Design

3.2.1 Factors

As mentioned in previous section, the major factor we investigate in this study is the appearance. The first step to design the study is to choose a proper way to quantify the factor under investigation. At this point, we try to be as general as possible in terms of measuring different aspects of the appearance. Therefore we only choose to parameterize the appearance as a categorical variable which has only a few of sampling levels. For example, we can use several textures to render the same geometric model. Each texture is a representative sample of its own class of textures. It also serves as a sampling point in the overall population of all appearance that can be used to render the same geometric model. Of course, the geometric model used in rendering can be another source of difference in appearance of an object. We do not explicitly distinguish between these two sources and only discuss the overall difference in appearance.

The complexity of a scene is another factor we want to investigate. Comparing to a simple scene, a complex scene may have more information for people to perceive and understand. The amount of information may influence the way people make judgment. For example, when asked question about a single object in the animation, people may focus on the single object to catch all visual details. However, when there are more objects moving at the same time, it is not easy for most people to notice all details. And the way they understand the visual information in order to answer a question can be different from before. In this study, we quantify the complexity of a scene by using different number of objects in animation.

Another parameter we also investigate in our study but not treating it as a factor is the Young's Modulus of deformable objects. A deformable object can have different stiffness parameterized by Young's Modulus. An object with very high Young's Modulus is very stiff that the appearance may not have as much influence on people's mind as it has on the same object but with a much lower Young's Modulus. Similarly, appearance may have limited effect on object that is extremely soft. We discuss the effect of appearance by looking at its influence on people's perception of an deformable object in a wide range of Young's Modulus levels. Thus we do not study Young's Modulus as a separate factor. More details on how we choose a proper range for Young's Modulus will be discussed in following chapters.

3.2.2 Stimuli Preparation

We use Blender [6] to create and edit models used in our study. All surface models are then passed to Tetgen [55] to generate the tetrahedral mesh needed for FEM simulation. We use barycentric interpolation to embed the secondary rendering mesh. Simulation and rendering models are adjusted so that most triangles in rendering models are very close to elements in simulation models if they are not contained by any element. This can effectively reduce the number of negative coefficients.

For deformation modeling, we use the VegaFEM [3] implementation of a co-rotational model [43] and an invertible finite element model [30]. We use an implicit backward Euler integrator because of their stability in practice. We use part of the Bullet physics library [11] for collision detection.

We use PovRay to render the deformed rendering mesh in each simulated frame. To suppress the interference of other factors that may influence people's perception of deformation, we choose a very simple simulation scenario: free falling of objects. The falling objects collide with the ground and bounce up or pile up. The ground is in green and the background is black for better observation. Videos are of the resolution 1024×768. They are displayed with a 23 inch screen.

We designed and displayed the script for both the pilot study and formal experiments using Psychopy [48].

3.2.3 Pilot Study

We first carry out a pilot study to collect primary data on people's perception of deformation. We choose to measure people's sensitivity to deformation in our animations. To do this, we use a yes-no experiment scheme. In the experiment, eight participants are asked to report whether they witness any deformation in the objects in each animation displayed. Each animation is 6 seconds long. Four simulation models are used: sphere, cylinder, cube and torus. We simulate each model with 32 different Young's Modulus levels ranging from $9e4$ to $8e7$. Thirteen appearances including seven shapes are then used to render these simulation results. We generate animations beforehand at fixed discrete Young's Modulus levels to gain better rendering quality. We then take advantage of the randomly interleaved staircase scheme [10, 35]. To determine the next stimuli to display for each appearance, we follow the method in interleaved staircase to compute the next stimuli level, but choose the closest Young's Modulus level among the 32 instead. This scheme converges to psychometric thresholds fast. For each appearance, we use Psignifit tools to fit a logistic psychometric curve for each participant [16]. Point of Subjective Equality (PSE), which represents stimuli parameter level corresponding to 50% probability that an object is perceived as deformable, is an important index in this result.

The result of our pilot study shows a great variance in people's PSE. However, people's answer on stimuli with Young's Modulus level near the two ends ($9e4$ and $8e7$) are very consistent. There are differences in PSE for different appearances in most people's data. Based on the pilot study, in our formal study, we choose a range for Young's Modulus that covers most people's PSE and extends to the ends where people's answer is consistent. Another observation from the study is that once they get used to the experimental scheme, people can give answers after viewing a small fraction of video. Thus shorter stimuli are used in the formal study.

3.3 Study I Experiments and Analysis

3.3.1 Experiment I.a: The Simulation of a Sphere Model

In our formal study, we quantitatively measure ***whether appearance can have a statistically significant effect on people's perception of deformation in animation of simple scenarios***. There is only one object in each stimulus. We choose to measure people's sensitivity to detectable differences of deformation. We use a fitted psychometric function and thresholds of Young's Modulus corresponding to detectable differences in deformation as a representation of people's perception of deformation. We intentionally include different high-level conceptual information and low-level visual details in the objects' appearances. Experiment I.a and I.b are performed separately. We further analyze ***whether high-level conceptual information or low-level visual details are dominant in the potential effect***.

3.3.1.1 The Design of Stimuli

We use the simulation of a sphere model in this study. To limit variables, appearance differs only in texture. Figure 3.3 shows the simulation and rendering meshes used in this experiment. The low resolution spherical model is converted to a tetrahedral mesh with 158 elements for simulation.

Seven appearances used in this study are given in Figure 3.4. Obviously, these appear-

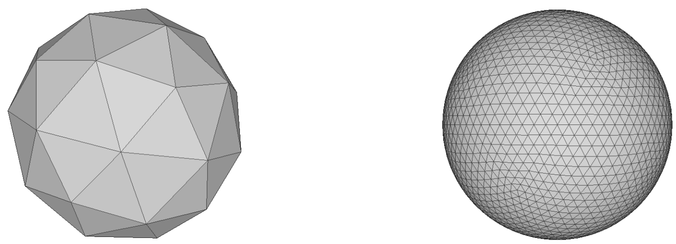


Figure 3.3: Spherical model used in simulation(left, 80 triangles) and the secondary model used for rendering(right, 5120 triangles).

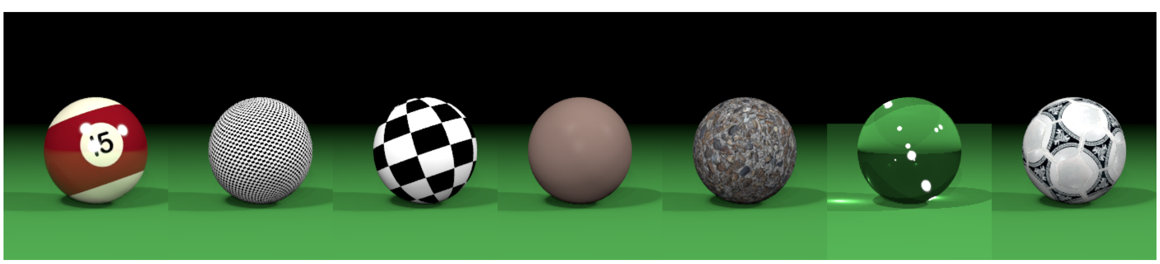


Figure 3.4: All appearances used to render physically based elastic deformation simulations of a sphere model. From left to right: Billiard, Checker-High, Checker-Low, Clay, Concrete, Refraction, Soccer.

ances have different amounts of low-level visual details as well as completely different high-level conceptual information. We use the 2 alternative forced choice (2AFC) protocol [5] which is bias free compared to a Yes-No scheme. Each stimulus is paired with an animation with the highest Young's Modulus as reference. The two stimuli are displayed successively in a random order. Participants are told that only one of each pair has rigid objects and they need to choose the one with deformable objects. Simulations at 9 Young's Modulus levels in the range $(9e4, 9e5)$ with about even intervals are used. The number of times these animations are displayed at these levels are 2,4,6,6,6,6,6,8 and 10. Animations with the softest object is displayed only twice because it can be easily distinguished from the hardest one. Each stimulus is 1 second which is long enough for participants to make a judgment. Sixteen people participated in this study(11M/5F), most of which are naive to physically based elastic deformation simulation and rendering. Their age ranged from 21 to 31. We explain each appearance to participants before the experiment. Full length example videos are then displayed to allow participants to observe each appearance. Then, the users begin watching stimuli.

3.3.1.2 Results and Analysis

For each appearance, we fit a logistic psychometric curve for each participant as well as for overall merged data.

Figure 3.5 shows the fitted curve of every appearance using overall merged data. In a 2AFC experiment, the lower asymptote of a fitted function is equal to 0.5, which is what one would get by guessing. So, we choose the Just Noticeable Difference (JND) threshold as a representation of people's perception of deformation. In this study, JND is a relative modification of Young's Modulus with respect to the most rigid level. JND corresponds to 75% of a chance that a stimulus is recognized correctly as deformable. This threshold indicates people's sensitivity to differences in perceived deformation. The point

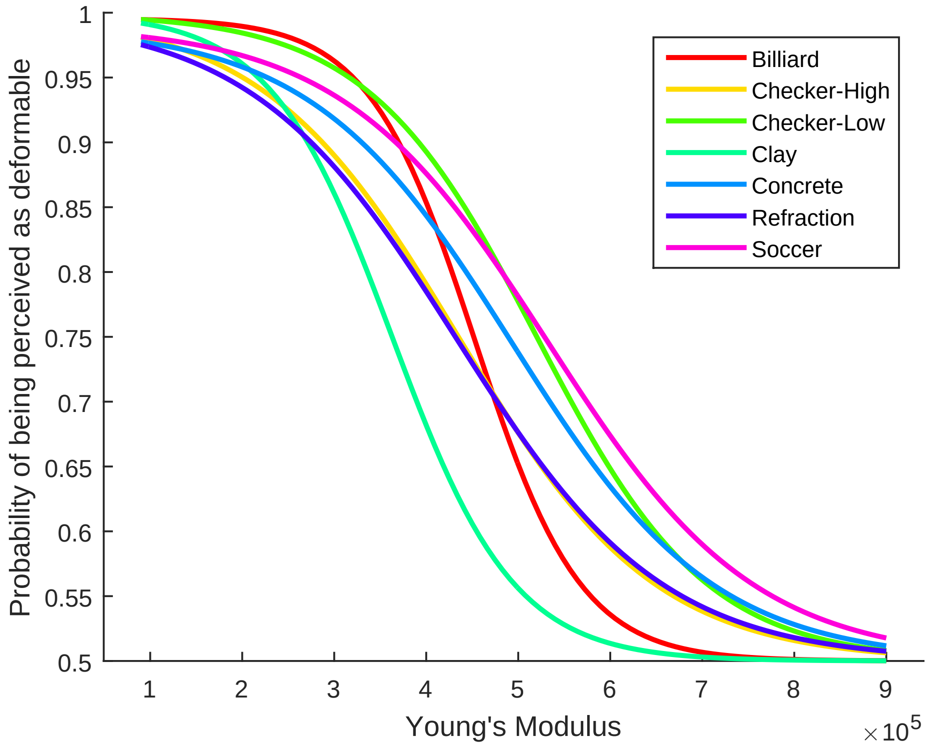


Figure 3.5: Psychometric curves fitted using accumulated data for every appearance in Experiment I.a.

and interval (95% confidence interval) estimation of JND for each appearance is listed in Table 3.1.

We further estimate the JND for each participant and each appearance. A one-way repeated measures ANOVA is performed. Appearance is found to have a significant effect on people's JND ($F_{6,90} = 2.9341, p < 0.015$). A Dunn-Sidak post hoc analysis shows that JND of appearance Clay is significantly different from the JND of appearance corresponding to the rightmost curves in Figure 3.5. This indicates that appearance can have a significant effect on JND in this study. Further investigation of curves in Figure 3.5 and estimation of JND for appearances in Table 3.1 shows a consistent result. Curves for different appearances diverge from each other in a wide probability range. Point estimation of JND for different appearances spread from 40% to nearly 60%.

Texture	JND	
	Point	Interval
Billiard	49.70%	-3.97% , +3.45%
CheckerBoardH	51.67%	-5.13% , +4.64%
CheckerBoardL	42.18%	-4.99% , +4.47%
Clay	59.67%	-3.67% , +3.71%
Concrete	45.14%	-5.78% , +5.51%
Refraction	51.91%	-5.14% , +4.99%
Soccer	40.81%	-2.22% , -10.67%

Table 3.1: Point estimation and interval estimation(95% confidence, relative to Point) of JND in Experiment I.a.

Although not every pair of JND for different appearances are significantly different, we can still tell the relative order of perceived stiffness of different appearances. An appearance with higher estimated JND generally corresponds to curves on the left in Figure 3.5. For such appearances, a larger reduction to Young's Modulus is necessary for people to perceive detectable deformation with the same accuracy as for other appearances. On the other hand, people have lower accuracy in detecting the difference of perceived deformation when a stimulus with such an appearance is displayed, compared to a stimulus with another appearance. In this sense, appearances with higher JND and curves on the left are perceived as less deformable. In contrast, lower JND indicates more deformable.

A careful investigation of Figure 3.5 leads to an interesting observation. Clay is perceived as the most rigid appearance. Billiard and Refraction (Glass-like) are all perceived as more deformable than Clay. Concrete is even closer to be perceived as the most deformable appearance. This order of perceived stiffness is not consistent with the order of stiffness simply inferred from the high-level information in these appearances. On the other hand, Checker-Low is perceived as one of the most deformable. Checker-High is closer to Clay. But neither of these two has clear high-level conceptual information since a sphere with such texture is not commonly seen in real life. All these lead to a positive correlation between the amount of low-level visual details and people's perceived softness. Soccer, Checker-Low and Concrete, which are perceived as more deformable, all have lots of high-contrast local features which are easy to track and recognize. In contrast, Clay, Refraction and Checker-High, either have plain color on the surface or have high frequency intensity variation that is hard for people to track. Comparing to texture with abundant low-level visual details, it can possibly be more difficult for people to recognize deformation from these textures. In this sense, the more easy-to-track low-level visual details an appearance has, the more deformable an object with such appearance is perceived to be. The effect of high-level information can be opposite to that of the low-level visual details in this study and may weaken the effect of each other. This can possibly explain why there is a single pair of significantly different textures. Even though, the order of these textures is more consistent with the effect of low-level visual details. In simple scenarios, the effect of low-level visual details is thus more dominant.

3.3.2 Experiment I.b: The Simulation of a Torus Model

3.3.2.1 The Design of Stimuli

To further investigate the effect of appearance we perform a second experiment using a torus model. In this experiment, we further consider visual information contributed by the rendering mesh. We use two secondary rendering meshes with similar bounding shapes but different structure and details to create four different appearances.

Figure 3.6 shows the triangular models we use for simulation and rendering. We use the left model in Figure 3.6 for simulation. It is a low resolution approximation of both rendering meshes. This model is further tessellated into 370 tetrahedral elements for simulation.

Four appearances used in this study are shown in Figure 3.7. We use the same exper-

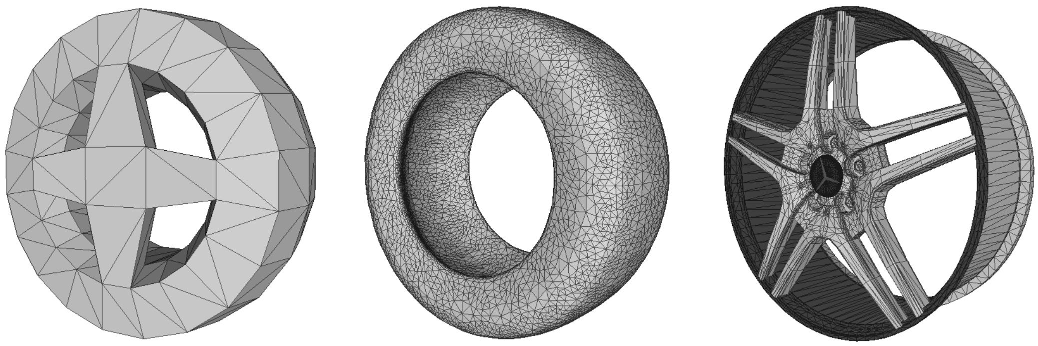


Figure 3.6: Torus model used in simulation(left, 310 triangles) and the secondary models used for rendering: tire(middle, 16416 triangles), rim(right, 11864 triangles).

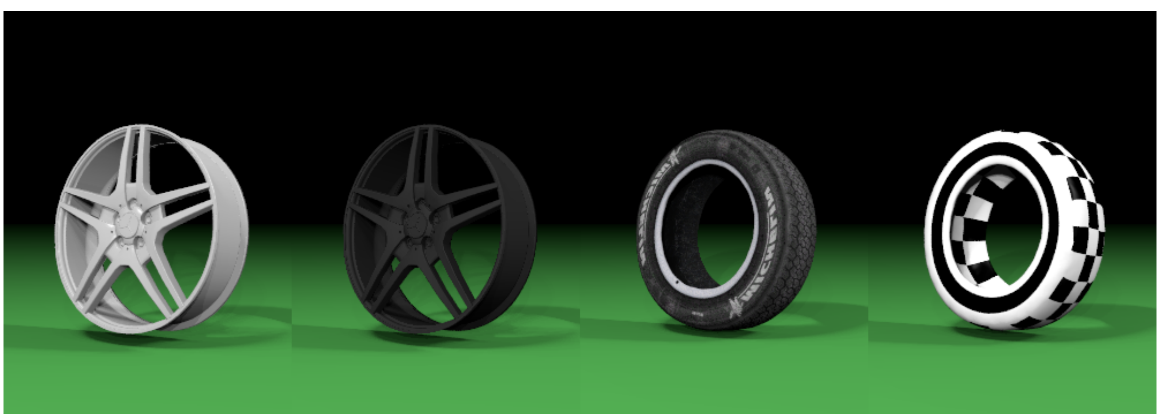


Figure 3.7: All appearance used to render physically based elastic deformation simulation of a torus model. From left to right: Rim-Bright, Rim-Dark, Tire-Black, Tire-Checker.

imental design as in previous experiment. We choose simulation at 10 Young's Modulus levels in the range $(1e5, 3.3e6)$. The number of times we display each of the four stimuli at the ten Young's Modulus levels are 2,4,6,6,6,6,6,6,8 and 10. Each stimulus is 2 seconds. We recruit another 11 volunteers (3F/8M) in this study. All of them are naive to the study. Their age ranges from 22 to 31.

3.3.2.2 Results and Analysis

Again, we fit logistic psychometric curves for each participant as well as for overall merged data. The fitted curve for each appearance using the merged data is shown in Figure 3.8.

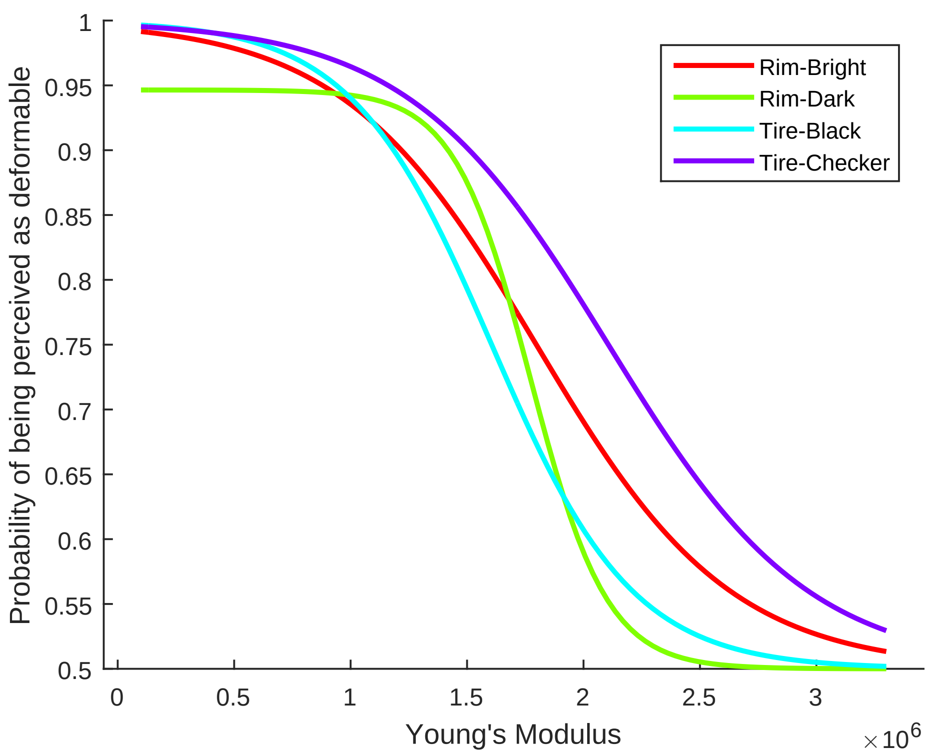


Figure 3.8: Psychometric curves fitted using accumulated data for all four appearances used in Experiment I.b.

Texture	JND	
	Point	Interval
Rim-Bright	45.53%	-6.59% , +5.28%
Rim-Dark	46.27%	-5.68% , +3.29%
Tire-Black	51.34%	-4.78% , +3.96%
Tire-Checker	36.16%	-7.44% , +6.44%

Table 3.2: Point estimation and interval estimation(95% confidence, relative to Point) of JND in Experiment I.b.

Estimation of JND for each appearance using overall data is reported in Table 3.2. One-way repeated measures ANOVA finds appearance to be a significant factor again ($F_{3,30} = 5.1647, p < 0.006$). Post hoc analysis finds the two tire appearances to have significantly different JND. The curves for the two tire appearances are on the opposite sides in Figure 3.8. The estimated JND for these two appearances have an approximate difference of 15%.

It is interesting to notice that Tire-Black, which should be the most deformable appearance according to the clear high-level information, is perceived as the most rigid one, even comparing to rim appearance. Tire-Checker has no clear high-level information. But it is perceived as more rigid than both rim appearance. Obviously, high-level information does not have significant influence here. Tire-Checker has the most easy-to-track and recognize low-level visual details. It is thus perceived as the most deformable one. Both rim appearances do not have colorful texture. But the complex rendering mesh causes more easy-to-track low-level visual details in rendered frames than Tire-Black appearance. Again, the amount of low-level visual details in appearance is positively correlated with the perceived softness. And this effect is dominant even if high-level information has opposite influence.

This study confirms the significant effect of low-level visual details in appearance. It further shows that low-level visual details can be from both texture and rendering mesh. Similarly, the amount of low-level visual details can also be controlled in many other ways. When designing the appearance of an object which will be simulated in 3D animation, visual reality is important. But if an artist wants to achieve a special effect by enhancing or reducing the perceived deformation, more attention may need to be paid to the design of low-level visual details.

3.3.3 Experiment I.c: The Effect of Complexity of Scenes

In the first two experiments, we statistically validate the significant effect of appearance in people's perception of deformation. Experimental data indicates that low-level visual details have a dominant effect while high-level information conceptual does not. However, people's perception of low-level visual details rely heavily on tracking of visual features. There has been study on the limit of people's ability in tracking multiple objects [61]. Complex scenes may contain too many low-level visual details which are beyond people's ability to recognize and process. We design our third experiment to study ***how complexity of a scene can influence the effect of appearance in people's perception of deformation***. We adjust the complexity of the scene by controlling the number of objects simulated. Furthermore, we measure people's perception of deformation by using their subjective rating of stiffness of objects. Although it may contain higher variance, it also provides more direct and abundant information.

3.3.3.1 The Design of Stimuli

We use the cylinder model in Figure 3.9 for simulation. It is tetrahedralized into 258 elements. To reduce the variance among people's rating, we display reference animation at the beginning and middle of the experiment. We use an extra appearance (Checker) in the reference animation. Only Gastank and Sausage are used as appearance in formal stimuli.

We use simulation at 7 Young's Modulus levels in the range $(9e4, 1.6e6)$. Five ref-

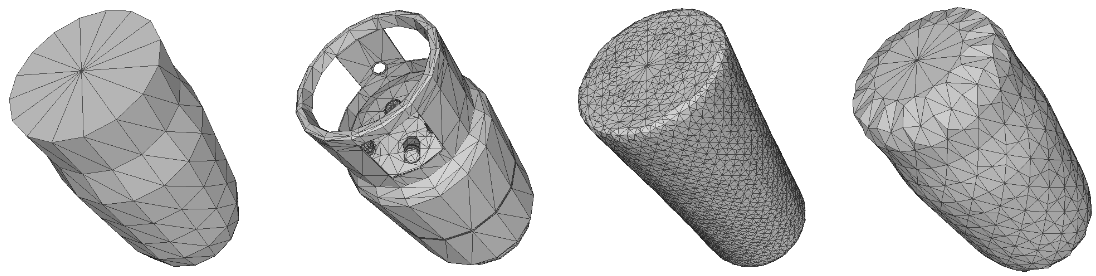


Figure 3.9: From left to right: cylinder model used in simulation (224 triangles) and the secondary rendering models gastank (7102 triangles), sausage (7680 triangles), checker (1344 triangles).

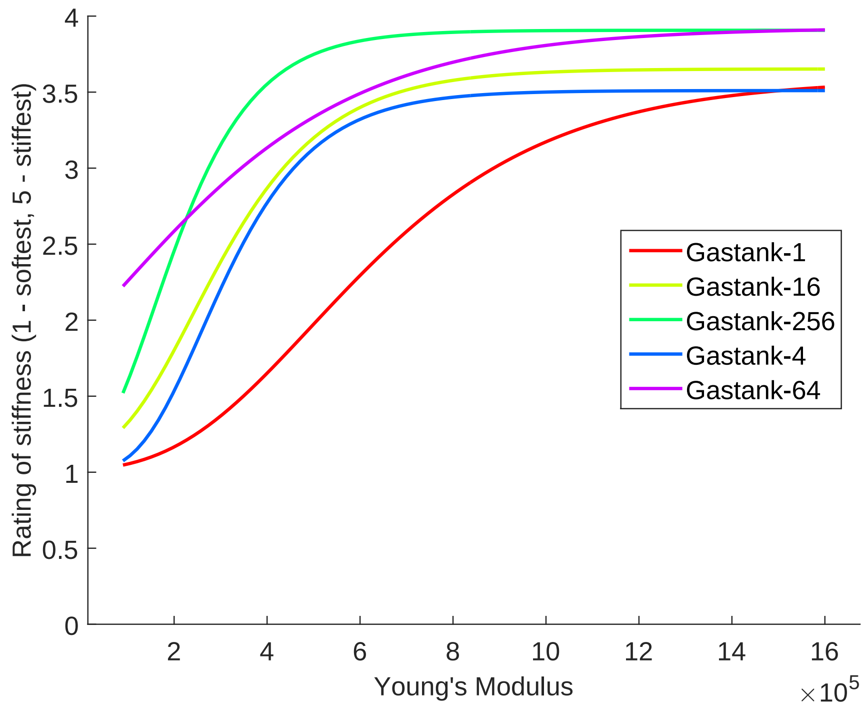
Figure 3.10: All appearance used in Experiment I.c. From left to right: Gastank, Sausage, Checker (only used in reference animation).

erence animations with different Young's Modulus levels evenly sampled in this range are displayed side-by-side for people to observe. All objects in these animations have a checker appearance. An integer number (1 to 5) is displayed on each animation to indicate its level of stiffness. Higher numbers correspond to stiffer objects. Reference animations are displayed in a loop until a participant is ready to proceed. For each appearance at each Young's Modulus level, we perform simulations of 1, 4, 16, 64 and 256 objects. Figure 1.2 shows example frames of stimuli with 64 objects. Overall, we have 70 distinct stimuli: 2 appearances $\times$ 5 Number of Objects (NoO) $\times$ 7 Young's Modulus levels (YM). Each 2-second stimulus is displayed four times in total. A random order is used for each participant to display the 280 stimuli. After watching each stimulus, participants are asked to rate the stiffness of objects by choosing a number from 1 (most deformable) to 5 (most rigid). We use the average of four ratings of each distinct stimuli as the final rating. We recruit 11 new volunteers (6M/5F) for this study. Their ages range from 23 to 40. All of them are naive to this study.

3.3.3.2 Results and Analysis

We first carry out a three way repeated measures ANOVA on the final rating. The result is reported in Table 3.3. To help understand the result, we also fit a function that represents the relationship between rating and Young's Modulus for each combination of appearance and NoO levels. These functions are shown in Figure 3.11.

One obvious and foreseeable result is the significant effect of YM in rating. We further apply the Dunn-Sidak post-hoc analysis on this main effect. We find, as we would expect, that all YM levels are significantly different from each other, and the rating mean increases as the YM level increases. In Figure 3.11, rating curves for all NoO and appearance combinations increase monotonically, which reflects the same effect. This indicates that people are sensitive enough to the difference between YM levels used in this study.



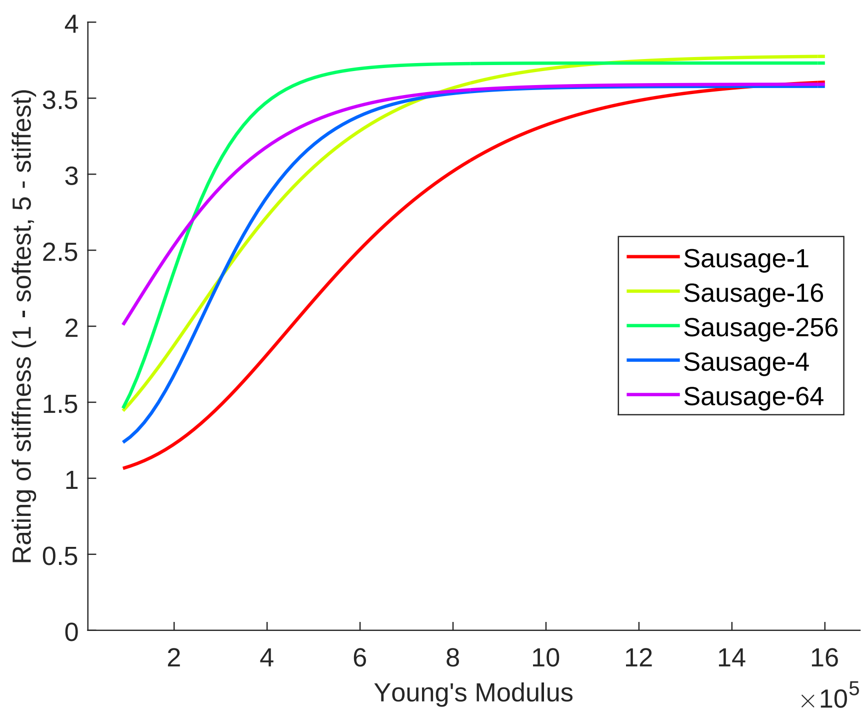


Figure 3.11: Fitted curves that represent the relationship between people's rating and Young's Modulus. Curves for objects with appearance Gastank(top). Curves for objects with appearance Sausage(bottom).

Factor	F-value	pValue	Post-hoc
NoO	$F_{4,40} = 44.6081$	$p < 4e - 14$	Higher rating for more objects
YM	$F_{6,60} = 170.0229$	$p < 1e - 35$	Higher rating for higher Young's Modulus
Appearance $\times$ NoO	$F_{4,40} = 2.735$	$p < 0.045$	Appearance has significant effect only for single object scenario
NoO $\times$ YM	$F_{24,240} = 8.0001$	$p < 1.2e - 19$	Number of Objects has more effect in Young's Modulus range (2e5, 8e5)

Table 3.3: Three way repeated measures ANOVA on people's rating of stiffness - 2 Textures $\times$ 5 Number of Objects (NoO) $\times$ 7 Young's Modulus levels (YM)

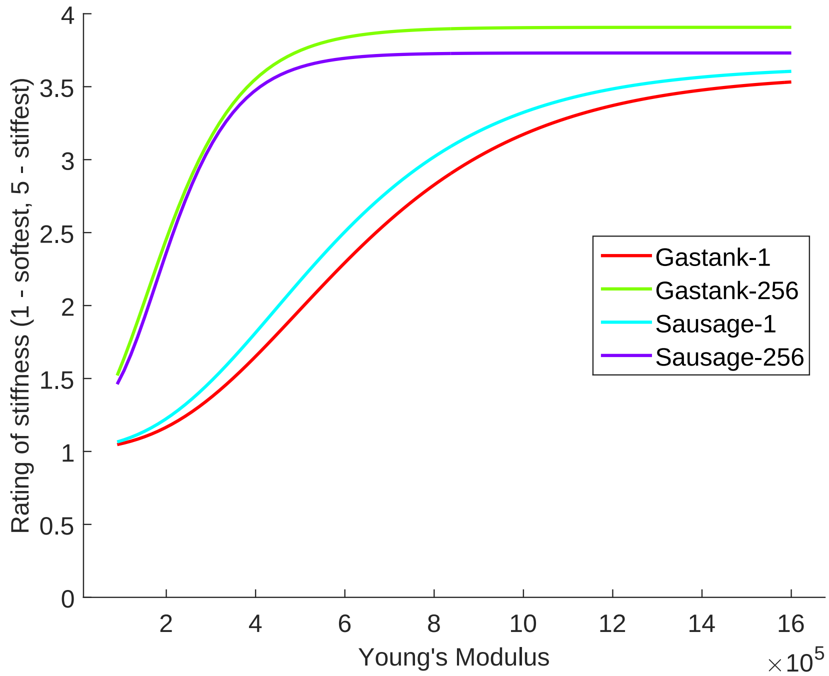


Figure 3.12: Curves for 1 and 256 objects with appearances Gastank and Sausage.

We do not find enough support for the significance of appearance as a main effect. However, the interaction between appearance and NoO is found to have significant effect ($F_{4,40} = 2.735, p < 0.045$). Post-hoc analysis indicates that mean ratings for Gastank and Sausage are significantly different only when NoO is 1. The curves for one Gastank object and one Sausage object in Figure 3.12 are diverged from each other throughout the entire YM range. The rating for Sausage is uniformly higher than Gastank when there is only one object. This result for a single object is consistent with our discovery in the previous two experiments. Although the high-level information in appearance Gastank hints for more rigid material, there are more low-level visual details in the form of high-contrast visual features in Gastank compared to Sausage. These features may allow the recognition of deformation to be easier. Thus Gastank can be perceived as more deformable. In simple scenario with one object, the effect of the low-level visual details obviously dominates the effect on people's rating again, even though the high-level information has opposite influence. This is not significant when there are more objects, however.

Another interesting observation is that NoO also has a significant effect on rating. Post-hoc analysis and fitted curves show that more objects generally cause higher rating for stiffness. In other words, as the number of objects under simulation increases, people begin to feel objects are less deformable, even though both the Young's Modulus and appearance of objects are not changed.

The study on people's limit in a multiple objects tracking task [61] can provide a potential explanation of this observation. As number of objects increase, the amount of low-level visual details increase accordingly. Tracking these cues and recognizing relative deformation is a more difficult task than simply tracking one object. Thus, the low-level visual details in a scene can simply overload people's ability to process them. When this happens, people are more likely to be distracted by overloaded details rather than processing them meaningfully. As a result, in complex scenes with lots of objects, people are less likely to recognize deformation and thus give higher rating. Experienced viewer may focus on a single object or local area in animation to avoid distraction. However, in entertainment scenarios where people are relaxed, they can easily be distracted by too many low-level visual details when watching unfamiliar scenes for a short period. This also explains why texture does not have significant effect when NoO is high. Low-level visual details in multiple Gastank or Sausage objects stimuli can both overload people's attention. Thus the effect of low-level visual details are depressed in both cases.

With more objects, there is also no significant effect that high-level information matters. However, in Figure 3.11, the functions for 64 and 256 Gastank are higher than the corresponding functions of Sausage in a wide range of YM. This may be a hint that as a perceptual system is overloaded, the high-level information indeed becomes important in perceived deformation. More study would be needed to confirm or refute this, though.

It is also important to notice that the effect of NoO is weaker when YM is closer to the two ends of selected range. The interactive effect of NoO and YM is partly caused by this. Intuitively, NoO has less influence on very soft and very rigid objects, which are more obviously soft or rigid.

3.4 Study I Conclusion

We perform three primary experiments to study the effect of objects' appearance on people's perception of deformation in computer generated animation. We validate that appearance can potentially have a significant effect on people's sensitivity to detectable differences of deformation and subjective rating of stiffness of objects. However, we find this effect to be significant only in simple scenarios with single object of moderate stiffness. Furthermore, the effect of low-level visual details in appearance is dominant comparing to high-level information in our studies. When low-level and high-level information contradict each other, low-level visual details will win out. We further investigate the influence of number of objects in the scene, controlling for the complexity of the scene. We discover that increasing the number of objects in stimuli can cause people to feel the objects are more rigid. This influence also depresses the effect of appearance on rating. Significant differences between ratings for one Gastank object and one Sausage object decrease as the number of objects in the stimuli increase. One possible explanation is that redundant low-level visual details in complex scenes could simply overload people's perceptual ability, thus causing more distraction than meaningful influence. This depresses the effect of low-level visual details as well as the overall effect of appearance. However, we have seen some hints that high-level information could actually assume a dominant role when large numbers of objects are simulated.

The results provide an important guide to artists in designing objects for animation. First, our quantitative validation proves that perception of deformability of an object can be adjusted by the design of its appearance. Second, our discovery shows that simply designing the appearance of an object as it is in real life may not be an optimal choice in controlling the perceived deformability. To enhance or reduce the perceived deformability, one can balance the design to include more or fewer low-level visual details in various forms (e.g. rendering mesh, texture, decoration, furs, etc.) or to achieve more realistic design. In complex scenes, however, more attention may need to be put on designing appearance that can deliver clearer high-level hints.

As a primary study, we only validate the significance of appearance, especially that of the low-level visual details, in people's perception of deformation. We qualitatively distinguish more and fewer low-level visual details in appearance while ignoring the source. Our three experiments are limited in the number of models and textures tested. More specifically designed experiments are necessary to verify our discovery in more general scenarios. Our following work will propose quantitative metrics of the amount and form of low-level visual details in influencing people's perception. We need to study typical aspects of low-level visual details which could be quantitatively measured (e.g. spatial-temporal frequency of intensity variation in texture). By removing the effect of high-level information, we can propose quantitative metrics which could be directly used by artists.

3.5 Study II Experiment Design

3.5.1 Factors

In this study, we investigate the low-level visual details as a sole factor in influencing people's perception of deformation. In order to remove the influence from high-level conceptual information in appearance, we choose the checkerboard texture which does not have clear high-level conceptual information. Also, we use the same sphere model in simulation to prepare all stimuli in this study. The difference in texture is thus the only source of difference in low-level visual details.

To quantitatively measure the effect of low-level visual details, we quantify the checkerboard texture by its spatial frequency and contrast. These two properties are thus treated as the actual factors under study which greatly determine the low-level visual details.

Again, we treat the complexity of scene as another factor to study. Because it shows to what extent the main factors under study can take effect.

3.5.2 Sample The Spatial Frequency-Contrast Domain to Create Textures

Based on previous studies of the sensitivity of the human visual system to rigid moving low-level visual details, we quantify the checkerboard pattern in two aspects: *spatial frequency* and *contrast*. As pointed out by Kelly [33], people's sensitivity to the same low-level visual details moving at different speed is different. In this section, we first describe how we determine a proper domain for spatial frequency and contrast according to the moving speed of object in our stimuli. People's sensitivity to low-level visual details in this domain varies significantly. We then measure people's perceived stiffness of objects textured using checkerboard pattern with spatial frequency and contrast samples uniformly distributed in this domain. By doing so, we can have an overall estimation of how perceived stiffness varies in this domain. A global model fitted from the sampling data can be useful in estimating people's perception of deformability for complex textures. It needs to be mentioned that we take advantage of previous studies on the Contrast Sensitivity Function (CSF) of a sinusoidal pattern only to determine a proper spatial frequency-contrast domain for sampling. The accurate position of the CSF, although may differ from that of a checkerboard pattern, does not influence much the domain chosen and our sampling strategy. To study the influence of low-level visual details, we vary the checkerboard pattern design to create different textures and measure the difference in people's perception of deformation when different textures are used. We referred to previous studies of the sensitivity of the human visual system to guide our design.

People's sensitivity to sine wave gratings of different spatial frequencies and temporal (flickering) frequencies was studied by measuring the threshold contrast for noticeable gratings [54]. The inverse of the measured threshold contrast was called the Contrast Sensitivity Function. The CSF at a fixed temporal frequency peaked for middle spatial frequencies (e.g. $3 \sim 10\ cycle/deg$) and fell rapidly in high spatial frequencies (and also in low spatial frequencies for low temporal frequency). Kelly measured and fitted the threshold surface for moving gratings [33]. Kelly found that the diagonal profiles of the surface, corresponding to constant velocities, were all of the same shape but only varied in the height and peak frequency. The measured surface was based on retinal velocity which differs from image plane velocity due to eye motion. Daly extended Kelly's model to include the three types of eye movements: natural drift, smooth pursuit, and saccade, so that image plane velocity could be used instead of retinal velocity [12]. The proposed model was:

$$G(\alpha, v) = c_0 k c_2 v_r (c_1 \alpha)^2 exp(-2c_1\alpha/\alpha_{max}) \tag{3.1}$$

$$k = 6.1 + 7.3|log(c_2 v_r/3)|^3 \tag{3.2}$$

$$\alpha_{max} = 45.9/(c_2 v_r + 2) \tag{3.3}$$

G is contrast sensitivity. $\alpha \equiv 2\pi(cycle/deg)$ is spatial frequency. c_0, c_1 and c_2 are chosen as $1.14, 0.67$ and 1.92 to accommodate screens with luminance levels $> 100\ cd/m^2$. v_r is measured in deg/s. v_r is retinal velocity and can be computed from image plane velocity v_i using the following formula:

$$v_r = max(0.15, v_i - min(v_{max}, g_{sp}v_i + v_{min})) \tag{3.4}$$

v_{min} and v_{max} are 0.15 and 80, which are natural drifting velocity of the eye and maximum smooth pursuit velocity. Following Daly, the so-called gain of the smooth pursuit eye movements g_{sp} is chosen as 0.82 [12]. Temporal frequency can be computed as $\omega = v_r\alpha$.

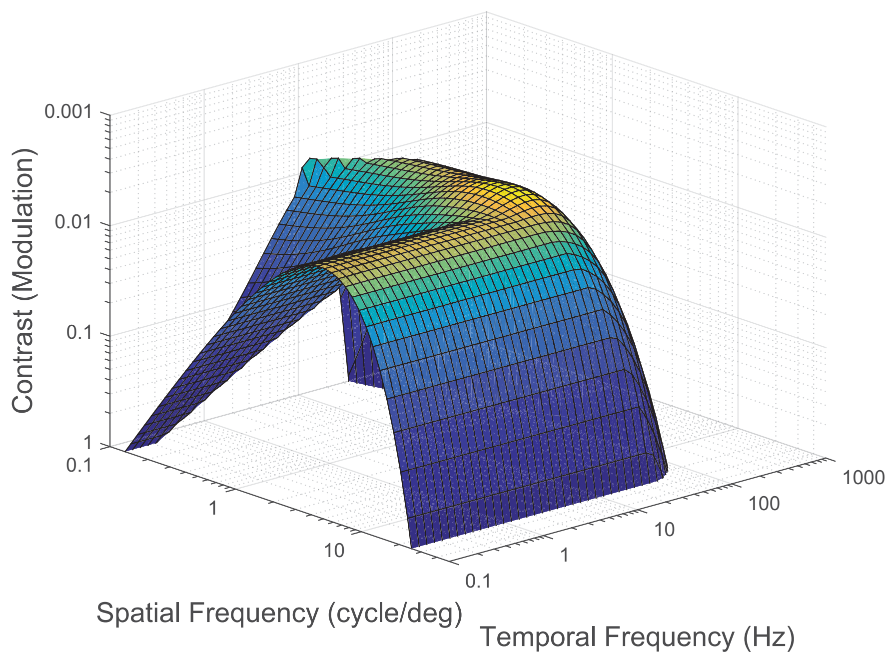
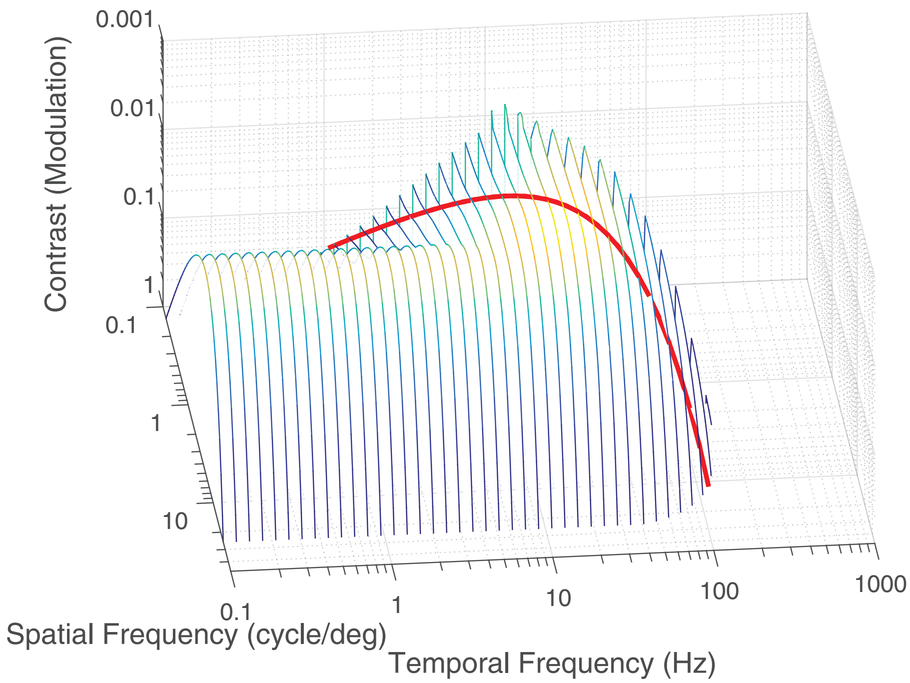


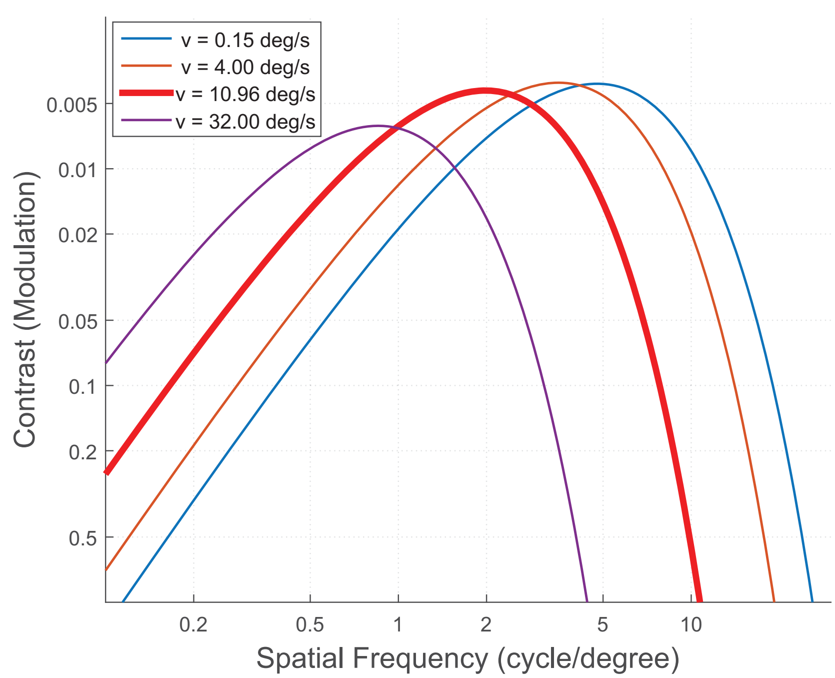
Figure 3.13: Contrast Sensitivity Surface

Figure 3.13 shows the fitted contrast sensitivity surface. Notice the spatial frequency, temporal frequency and contrast sensitivity axis are given in a log scale in all figures in this paper. In log space, a diagonal profile of the surface as shown in Figure 3.14a corresponds to a CSF measured at that constant velocity.

Figure 3.14b shows the projection of the CSF corresponding to several different velocities onto the spatial frequency-contrast plane. As velocity increases, the CSF shifts in the low spatial frequency direction.



(a) A 45° profile on the surface corresponds to constant velocity



(b) CSF at constant velocities (The red thick curve is the projection of the red profile in Figure 3.14a)

Figure 3.14: Diagonal profile (top) and Its projection on the Spatial Frequency-Contrast plane (bottom)

In our study, stimuli only differ in the object's texture and stiffness. The velocity of the object, which is also the velocity of the texture, is the same in all stimuli ($10.96\ deg/s$ at the point of collision, after which there may be small variations). We thus consider CSF for texture moving at speed $10.96\ deg/s$ (the red thick curve in Figure 3.14b). As a result, we discretize the *spatial frequency-contrast* $log - log$ domain by regularly sampling the entire region that the red curve lies in. We create a texture for each sampling point using the corresponding spatial frequency and contrast.

It needs to be mentioned that although such a design is intuitive and can reflect the relative difference in low-level details in texture, it is only based on a quantitative model of the rigid texture, rather than the deforming texture. A discretization and sampling of the domain in a model that measures the deforming texture might be more representative, but is left as a topic for future study.

3.5.3 Stimuli Preparation

Similar to previous study, we choose the Finite Element Method (FEM) based methods for simulation and adjust the Young's Modulus to achieve different stiffness. We use the same framework for simulation. We use Blender and Tetgen to generate the tetrahedral mesh needed for FEM simulation. We use a sphere model with 158 tetrahedral elements for simulation. A higher resolution rendering mesh with 5120 triangles is then used for rendering. We use PovRay to render the deformed rendering mesh in each simulated frame.

We simulate the simplest scenario, a free-falling sphere, to generate stimuli. In all stimuli, a sphere falls, collides with the ground and bounces up. Figure 3.15 shows examples of our stimuli used in this study. All videos are of resolution 1024×768. They are displayed with a 23 inch screen with average luminance $> 100cd/m^2$. We designed and displayed the script using Psychopy.

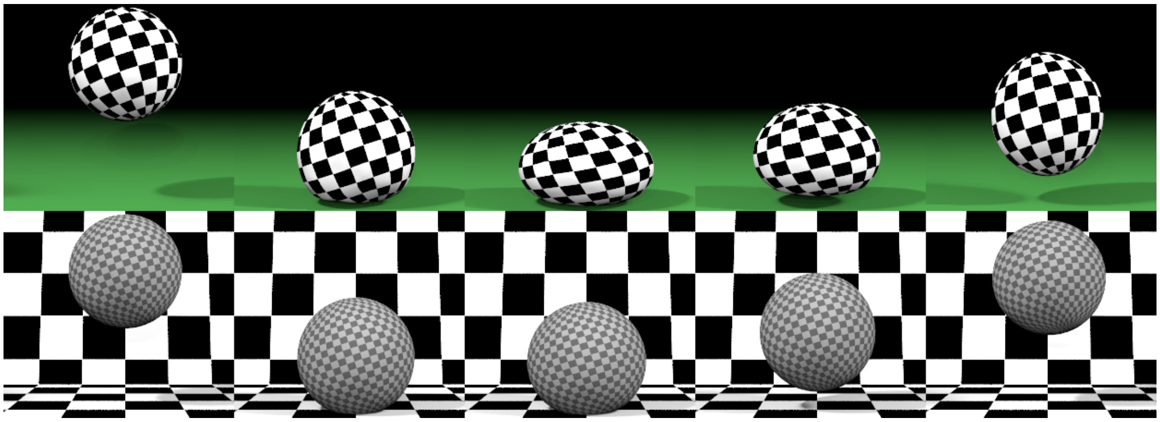


Figure 3.15: Frames taken from two stimuli. They are rendered from a physically based simulation of deformable sphere models with differing Young's Modulus. We combine different levels of contrast and spatial frequency to create 25 different checkerboard textures, two of which are shown above. We further study how backgrounds can influence the effect of low-level visual details using stimuli like the lower one.

3.6 Study II Experiments and Analysis

We perform two experiments in this study. In the first experiment, we render the background as green and black to allow people to focus on the appearance of the foreground object. In the second experiment, we add checkerboard background to increase the complexity of the scene and study the influence of that.

3.6.1 Experiment II.a: The Effect of The Checkerboard Pattern

In this experiment, we study whether and how checkerboard textures on deformable objects with different spatial frequency and contrast can influence people's perceived stiffness. We sample spatial frequency and contrast each at five levels to create 25 different checkerboard patterns. Stimuli are created by simulating spheres of different Young's Modulus, and for each of these rendering them with these textures. We then carry out statistical analysis on participants' rating of stiffness of these stimuli to reveal any significant pattern. As mentioned previously, we use the CSF of a sinusoidal pattern as an approximation of the CSF of a checkerboard pattern. This is mainly to help understanding the result. The exact position of CSF is not used in computation.

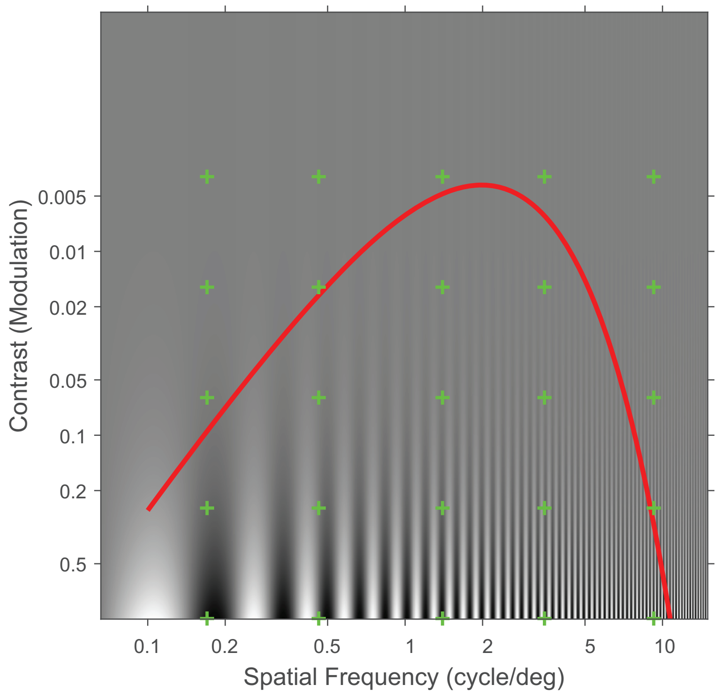


Figure 3.16: Sampling on the *spatial frequency-contrast* plane. Green crosses represent sample positions. The red curve is the CSF at constant velocity $10.96\ deg/s$. Spatial frequency values are as seen from a particular fixed distance.

3.6.1.1 The Design of Stimuli

Stimuli on screen are $27.5\ inches$ from participants. The viewing angle of the sphere is $3.25°$. The viewing angle of the entire falling path is $5.84°$. The image plane velocity of the falling sphere right before collision is $10.96°/s$. The CSF corresponding to this velocity is plotted in Figure 3.14a and 3.14b in red. Correspondingly, we adjust the size of the color block in the checkerboard texture such that the texture on the sphere in the final animation has spatial frequency spreading in a range wide enough to cover the CSF. The sinusoidal CSF corresponding to this velocity is plotted again in Figure 3.16 in red. Each point in this domain corresponds to a texture with specific frequency and contrast. The area below the CSF curve corresponds to rigid textures that people are more sensitive to perceiving. In Figure 3.16, a reference grating pattern is rendered as background for illustration throughout the *spatial frequency-contrast* domain. We want to examine the area where the CSF lies when sampling. Our design thus samples the spatial frequency at five levels: $0.17, 0.46, 1.39, 3.47, 9.26\ cycle/deg$. To sample the contrast domain, we choose five contrast levels: $1/1, 1/4, 1/16, 1/64, 1/255$. We combine each pair of samples in spatial frequency and contrast to form a grid-like sampling in the domain as shown in Figure 3.16. The varying grating pattern in background of Figure 3.16 is only for illustration purpose.

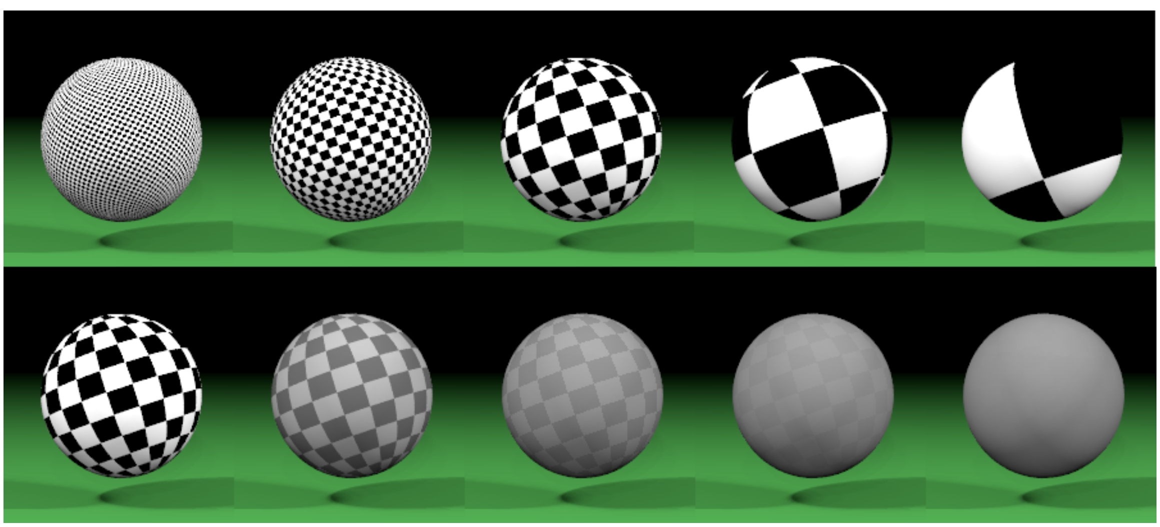


Figure 3.17: Example frame from animation with 5 different spatial frequency levels (top) and 5 different contrast levels (bottom)

The 25 samples in the domain encompass almost the entire region of the CSF. The background is rendered in black and green for simplicity. Some example frames are shown in Figure 3.17. Although the texture near the silhouette of the sphere in the animation contracts due to the projection, the spatial frequency of the texture is relatively uniform in the center area of the sphere. Also, the sphere does not rotate much during the short simulation. As a result, the texture on the sphere may only translate and deform due to simulation.

We want to measure the influence of low-level visual details at different stiffness levels. So, we also sample the Young's Modulus domain uniformly at 7 different levels in the range $5e4 \sim 1.2e6$. The 25 checkerboard textures are used to render simulation at each of the 7 Young's Modulus levels. Overall, we have 175 distinct stimuli: 5 Spatial Frequencies (SF) $\times$ 5 Contrast $\times$ 7 Young's Modulus (YM) levels. Each stimulus is 1 second long which we have found to be sufficient for viewers to evaluate in practice.

Participants are asked to rate the stiffness of objects using an integer from 1 to 5 after watching each stimulus. 5 represents the most rigid and 1 most deformable. Each distinct stimulus is displayed three times in the experiment and their average is used as the final rating for that stimulus. All stimuli are displayed in a random order. Before the actual study, five reference stimuli (with Young's Modulus $5e4, 9e4, 2e5, 6e5, 1e6$) are displayed side-by-side with integers labels from 1 to 5 indicating stiffness. These reference stimuli are displayed in a loop during several short breaks in the experiment. Participants are told that the stiffness and texture of each object can vary and they need to rate according to their observation. We recruited 13 participants who are naive to this study. Their ages range from 20 to 40.

3.6.1.2 Results and Analysis

We first perform a three way repeated measures ANOVA on participants' rating data. Significant factors ($pValue < 0.05$) are reported in Table 3.4. Our studies meet Mauchly's

Factor	F-value	p-Value	Post-hoc
SF	$F_{4,48} = 9.29$	$p < 1.21e-5$	Lower rating for middle frequency range ($0.5 \sim 1.4$ cycle/deg)
Contrast	$F_{4,48} = 15.22$	$p < 4.17e-8$	Lower rating for higher Contrast
YM	$F_{6,72} = 577.72$	$p < 8.65e-59$	Higher rating for higher Young's Modulus
SF × Contrast	$F_{16,192} = 5.29$	$p < 3.59e-9$	Lower rating for higher Contrast in middle frequency range ($0.5 \sim 1.4$ cycle/deg)
SF × YM†	$F_{24,288} = 3.13$	$p < 2.86e-6$	Larger effect of spatial frequency in Young's Modulus range ($4e5 \sim 8e5$)
Contrast × YM†	$F_{24,288} = 3.77$	$p < 3.42e-8$	Larger effect of contrast for high Young's Modulus range ($> 4e5$)

† Only Lower-Bound adjustment reports p-value higher than 0.05.

Table 3.4: Three way repeated measures ANOVA on people's rating of stiffness of a free-falling sphere in plain background - 5 Spatial Frequency (SF) Levels × 5 Contrast Levels × 7 Young's Modulus (YM) levels

test for sphericity, meaning it is valid to use the standard p-value. Despite this, we also take the Greenhouse-Geisser, Huynh-Feldt, and Lower Bound adjustments of the p-value into consideration.

The first observation from the ANOVA results is that ***(1) the checkerboard texture, and thus the low-level visual details, can significantly influence people's rating of stiffness on a plain background***. This claim is supported by the fact that spatial frequency, contrast and their combination are all found to be significant factors. A Dunn-Sidak post-hoc analysis further explores how people's rating is influenced by contrast and spatial frequency. People's rating of stiffness is lower for spheres with textures of higher contrast. The significant difference ($p < 0.007$) between average rating at the highest and lowest contrast is close to 0.3. On the other hand, people's rating is lower for mid-range spatial frequency ($0.5 \sim 1.4$ cycle/deg). An investigation of the post-hoc analysis on the combination of spatial frequency and contrast is consistent with the above results. All these lead to our second observation: ***(2) checkerboard textures with certain spatial frequency and contrast combinations can reduce the perceived stiffness of spheres.*** For these spatial frequency-contrast combinations, all of which are well below the CSF curve, people perceive the objects to be more deformable. A quick analysis on the combination of either spatial frequency or contrast with Young's Modulus shows that ***(3) the effect of spatial frequency and contrast, and thus of low-level visual details, is larger for higher Young's Modulus.*** In other words, the effect of low-level visual details can be larger in rendering stiffer objects.

To visualize the effect of spatial frequency and contrast in an intuitive way, we plot the isosurfaces of people's rating using the average rating of all participants at every factor level. Figure 3.18 shows the isosurfaces. Clearly, isosurfaces of higher rating correspond to higher Young's Modulus values. However, there is an obvious bump in the high contrast and middle spatial frequency area. A bump indicates that a sphere with texture in this area

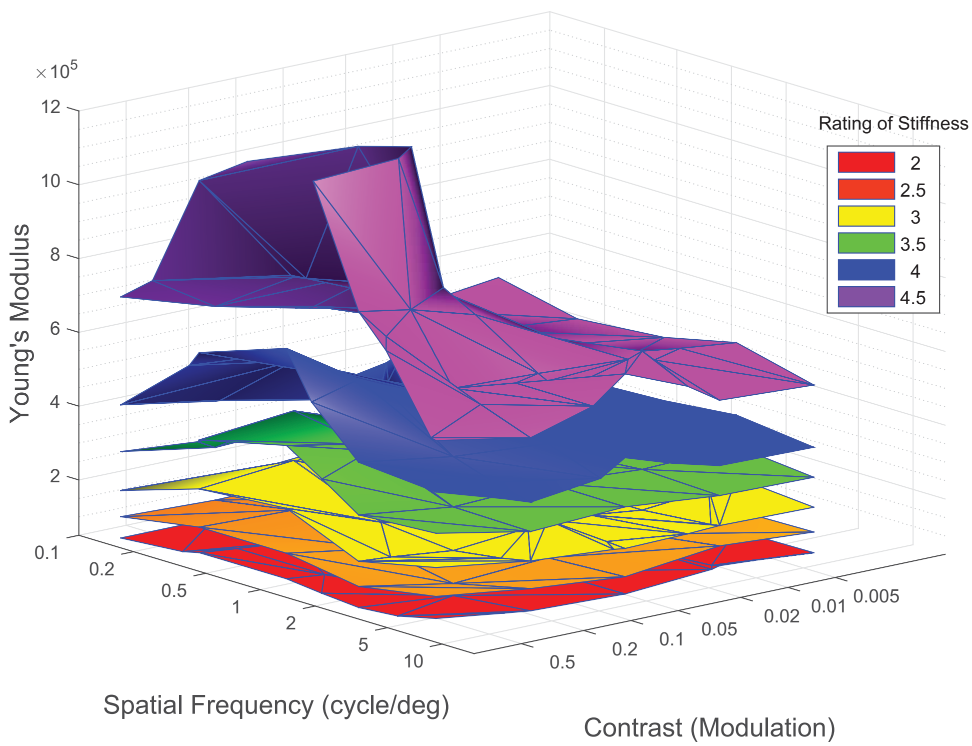


Figure 3.18: Isosurfaces of participants' rating in Experiment II.a. See supplementary material for 360 ° view of this figure.

but a higher Young's Modulus can be rated the same as a sphere with a different texture but a lower Young's Modulus. In other words, a sphere with such texture will be perceived as softer than a sphere with a different texture but the same Young's Modulus. This explains our second observation more intuitively. Meanwhile, the bump is higher for isosurfaces of higher rating, which supports observation (3) that the effects of low-level visual details are larger for stiffer objects.

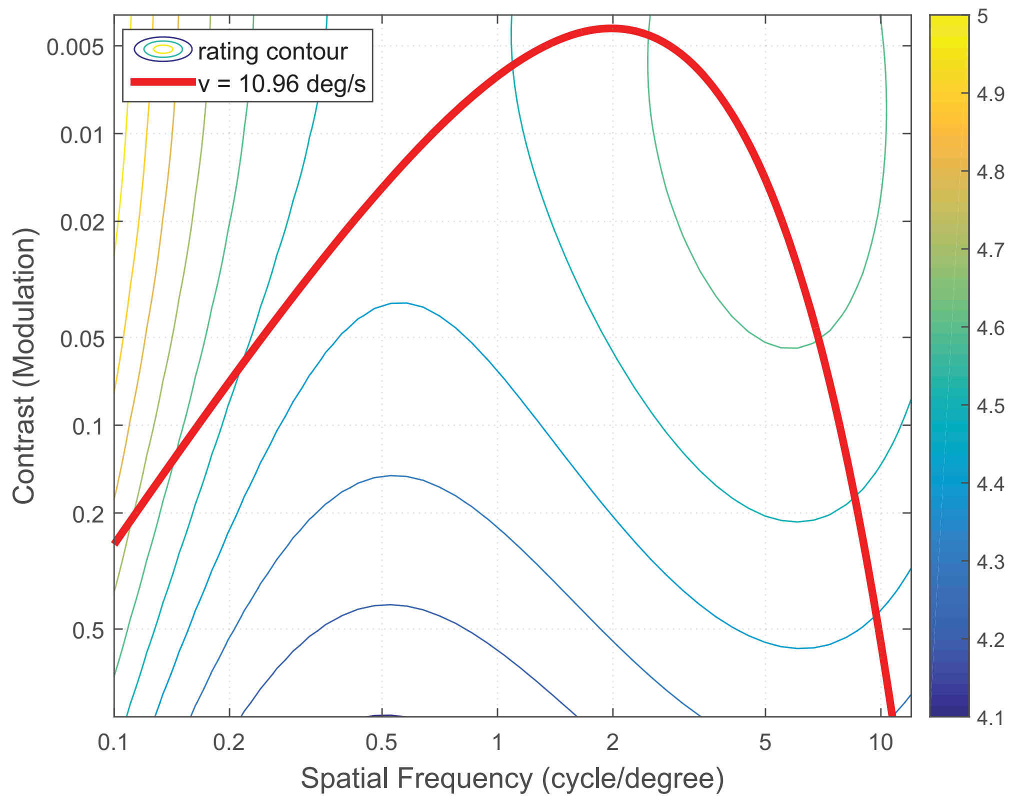


Figure 3.19: Contour of rating at Young's Modulus level $6e5$ and CSF at constant velocity $10.96\ deg/s$ in Experiment II.a.

To further investigate the second observation, we fit a generalized linear model from people's rating. We then plot the contour of people's rating as a function of spatial frequency and contrast at a fixed Young's Modulus level (e.g. around $6e5$). Figure 3.19 shows the contour as well as a constant velocity CSF of a sinusoidal pattern for reference.

As can be seen in Figure 3.19, the area with high contrast (around 0.5) and middle frequency (around $0.5 cycle/deg$) has a rating much lower than other areas. The difference is more than 0.5 between this area and most other areas. A much smaller deviation of rating can be found in other areas (less than 0.3). It is also interesting to notice that the CSF (the red curve) envelops the low rating area, although there is some discrepancy in their peak value (i.e. the CSF is similar to the contour in shape, but shifted slightly toward higher spatial frequency). This is again consistent with our second observation. The area under the CSF corresponds to stimuli that people are able to discern, which is also the area to which people give lower ratings of stiffness.

Again, the sinusoidal CSF in Figure 3.19 could deviate from a checkerboard CSF. Thus Figure 3.19 only qualitatively illustrates the relationship between the contour and the CSF. Also, the CSF is limited in quantitatively differentiating people's sensitivity to different stimuli. It seems reasonable to expect that our second observation also holds for a checkerboard CSF. The similar shape and position of the contour and the CSF hints that there might be a stronger relationship between people's sensitivity and people's rating of stiffness. However, more study is necessary to determine whether such a relationship actually exists, and if so, what the relationship is.

3.6.2 Experiment II.b: The Effect of Complexity of Scenes

In this study, we investigate whether a complex background can influence the effect of low-level visual details. Basically, we perform an experiment similar to Experiment II.a but using a checkerboard background. We first carry out statistical analysis of this study alone. We then analyze data from both studies to further investigate the effect of complex backgrounds.

3.6.2.1 The Design of Stimuli

We sample the spatial frequency, contrast, and Young's Modulus domain for the dropping sphere in the same way as in Experiment II.a. To create stimuli used in this study, we add a vertical plane in the animation. The checkerboard texture with spatial frequency $0.46\ cycle/deg$ and highest contrast is then mapped to the ground and vertical plane. People are very sensitive to this spatial frequency-contrast combination. However, projection does cause some contraction of the pattern on the ground. The experiment script is the same as Experiment II.a. We use the same reference stimuli without background to avoid bias caused by reference. Another 13 people participated in this study. They were all newly recruited and naive to our study. Their ages range from 19 to 30.

3.6.2.2 Results and Analysis

We again perform a three way repeated measures ANOVA on rating data. Significant factors ($pValue < 0.05$) are reported in Table 3.5. Mauchly's test again shows no violation of sphericity. One can easily find that the contrast and the combination of contrast and spatial frequency are no longer significant factors. Spatial frequency is still reported as a significant factor. But Dunn-Sidak post-hoc analysis reveals a different pattern of influence from spatial frequency. Only ratings at spatial frequency level $3.47\ cycle/deg$ are found to be significantly higher than ratings at other spatial frequencies. The effect is found to be larger when Young's Modulus is in the range ($4e5 \sim 8e5$). In short, ***(4) the checkerboard background reduces the effect of low-level visual details on foreground objects in influencing people's perception of deformation.***

The reduced effect of foreground texture can be seen more intuitively in Figure 3.20. The bump on isosurfaces in Figure 3.18 disappears. Instead, there are valleys on most isosurfaces at spatial frequency $3.47\ cycle/deg$. The isosurface of high rating (e.g. 4.5)

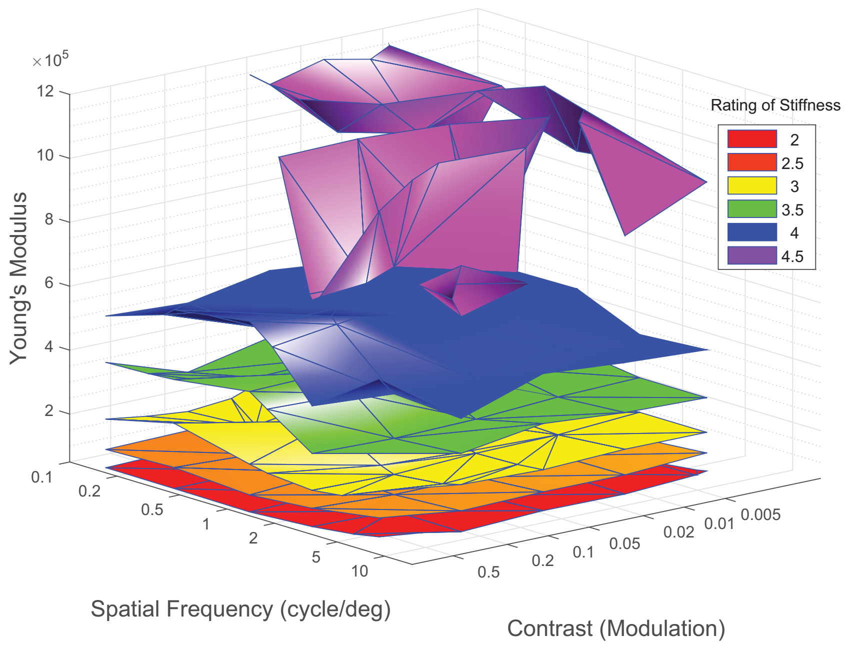


Figure 3.20: Isosurfaces of participants' rating in Experiment II.b. See supplementary material for 360 ° view of this figure.

Factor	F-value	p-Value	Post-hoc
SF	$F_{4,48} = 5.9396$	$p < 0.0006$	Significant higher rating for specific SF level ($3.47 cycle/deg$)
YM	$F_{6,72} = 349.3260$	$p < 3.84e-51$	Higher rating for higher Young's Modulus
SF × YM[†]	$F_{24,288} = 2.0129$	$p < 0.004$	Larger effect of SF in Young's Modulus range ($4e5 \sim 8e5$)
Contrast × YM[†]	$F_{24,288} = 1.8471$	$p < 0.01$	Small difference at random contrast × YM combinations

[†] Both Greenhouse-Geisser and Lower Bound adjustment report p value greater than 0.05.

Table 3.5: Three way repeated measures ANOVA on people's rating of stiffness of a free-falling sphere in checkerboard background - 5 Spatial Frequency (SF) Levels × 5 Contrast Levels × 7 Young's Modulus (YM) levels

is noisy, but a valley is still visible at spatial frequency $3.47\ cycle/deg$. There is almost no slope along the contrast axis in any isosurface which means that contrast does not have significant influence.

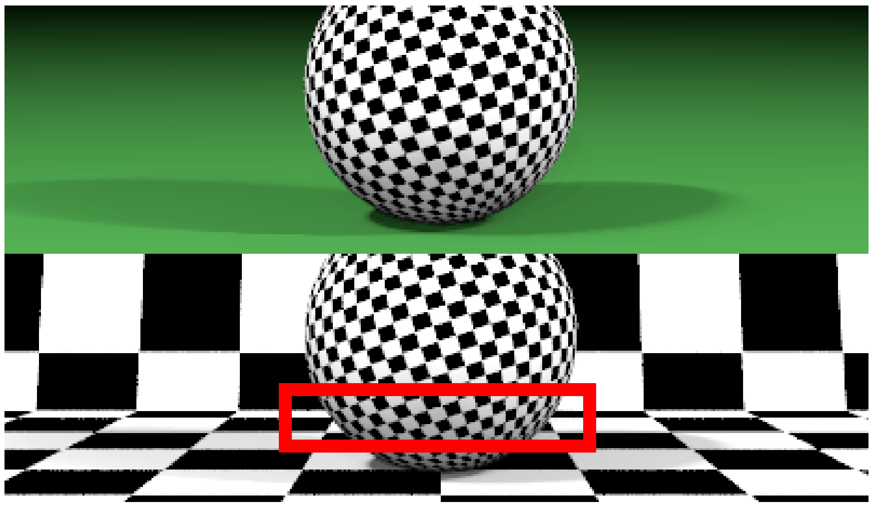


Figure 3.21: Frames from stimuli with spatial frequency $3.47 cycle/deg$, without background (top) and with background (bottom)

To investigate the reason for the high rating for spatial frequency $3.47\ cycle/deg$, two frames from the stimuli with texture of this spatial frequency are compared side by side in Figure 3.21. As can be seen in the area marked by a red rectangle, the contracted checkerboard background near the bottom of the sphere has a similar vertical spatial frequency as the texture on the sphere. This can possibly interfere with people's judgment of deformation.

We also fit a generalized linear model for rating in this study. The contour of the rating model as a function of spatial frequency and contrast at Young's Modulus level $6e5$ is shown in Figure 3.22. Noticeably, the range of rating these contours represent is only 0.35 comparing to nearly 1 in Figure 3.19, which means the effect of spatial frequency and

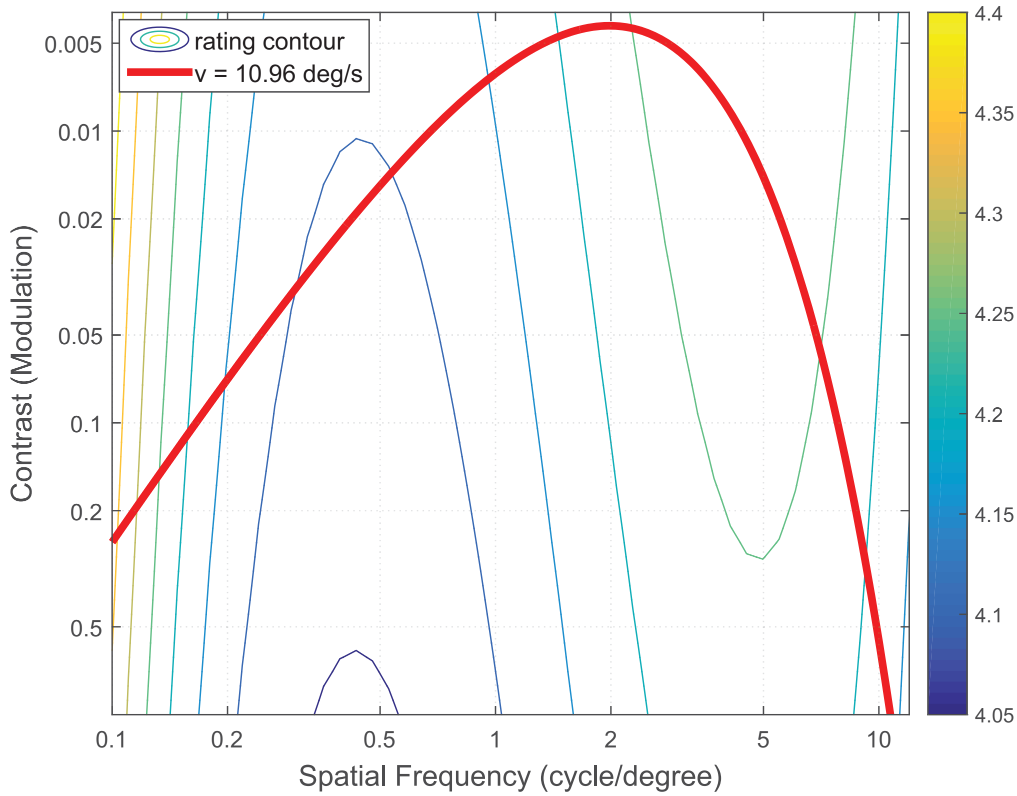


Figure 3.22: Contour of rating at Young's Modulus level $6e5$ and CSF at constant velocity $10.96 deg/s$ in Experiment II.b.

contrast on people's rating is reduced when a complex background is present. On the other hand, the contour lines are almost vertical which indicates less variance along the contrast axis, thus no significant effect of contrast.

We take advantage of the four way repeated measures ANOVA using data from both studies to investigate how a background can reduce the effect of foreground texture. The analysis is similar to previous ANOVA analysis but with an additional between-subject factor: W/O Background (BKGD). The result is shown in Table 3.6.

An interesting result is that there is not enough evidence to claim BKGD as a significant factor. In other words, BKGD does not uniformly reduce or increase people's rating of stiffness. By investigating the combined effect of BKGD with contrast or with spatial frequency and contrast, we discover that ***(5) a complex checkerboard background can reduce ratings for spheres with checkerboard textures that correspond to areas close to or above the CSF in the spatial frequency-contrast domain***. For spheres with checkerboard textures that are difficult for people to discern, people may tend to observe the silhouette of the sphere for any deformation. The background texture in this case can possibly provide low-level visual details near the silhouette of a sphere, thus allowing people to make a better judgment. Since such extra information has nothing to do with foreground texture, it can always be used by participants whenever they are not sensitive to the foreground texture. Thus, the effect of low-level visual details in the foreground is reduced by the low-level visual details in the background.

3.7 Study II Conclusion

We have performed two subjective rating experiments to study how low-level visual details can influence people's perception of physically simulated deformation. Participants rate animation of a free-falling and bouncing spheres with different Young's Modulus and checkerboard textures. Ratings of stiffness of spheres are given to each stimulus. We

Factor	F-value	p-Value	Post-hoc
BKGD*	$F_{1,24} = 2.1145$	$p < 0.16$	-
BKGD × Contrast	$F_{4,48} = 15.2245$	$p < 4.17e-8$	Background reduces rating more for lower contrast
BKGD × YM†	$F_{6,72} = 577.7164$	$p < 8.65e-59$	Background reduces rating more for higher Young's Modulus
BKGD × SF × Contrast†	$F_{16,192} = 5.2852$	$p < 3.59e-9$	Background reduces rating for specific SF × Contrast combinations, which people are less sensitive to.

* Repeated measure ANOVA does not find enough support for significance. Listed here just for reference.
† Only Lower-Bound adjustment reports p-value higher than 0.05.

Table 3.6: Four way repeated measures ANOVA on people's rating of stiffness using data from both studies - Within subject factors: 5 Spatial Frequency (SF) Levels × 5 Contrast Levels × 7 Young's Modulus (YM) levels - Between subject factors: 2 W/O Background (BKGD) Levels. Only factors including BKGD are listed here.

design the textures by using spatial frequency and contrast uniformly sampled from their $log - log$ domain.

A statistical analysis proves the significant effect of contrast and spatial frequency of the checkerboard texture on people's rating of stiffness in front of a simple background. It further shows that a checkerboard texture with certain combinations of spatial frequency and contrast can reduce the perceived stiffness. This effect is larger for stiffer spheres. However, the effect of a checkerboard texture on the sphere is much weaker in animation with a complex background (e.g. a high contrast checkerboard background). Only the rating of textures with a specific spatial frequency is significantly higher than others. This may be due to the interference on the human visual system of the background texture with similar spatial frequency. A combined analysis using data from both studies shows that a high contrast checkerboard background can significantly reduce people's rating of stiffness for spheres with textures that are difficult for people to discern.

Our results provide support for the importance of low-level visual details in design of textures that aim to enhance or reduce people's perception of deformation. This includes the low-level visual details in both foreground objects and background. For simple backgrounds, using textures that people are able to discern on an object can reduce the perceived stiffness. Using high contrast textures in the background can reduce the perceived stiffness of objects with textures that are hard for people to discern but will also reduce the effect of foreground low-level details.

A natural extension of our discovery would be to combine our result with wavelet transformation or DFT to provide a metric for measuring the perceived stiffness of a complex texture. Another area for future work is to propose a metric that can quantitatively measure the rate of change of low-level visual details in a deforming texture, which can be used as a more natural and clear model of visual stimuli with deformation. On the other hand, we may perform a more systematic study on the effect of background texture. Even from our limited study, we can tell that background can influence the foreground texture's effect on people's perception of deformation. We can also consider investigating other sources of low-level visual details like the shape of the rendering mesh.

4. FACTORS IN THE SIMULATION OF ANIMATIONS WITH DEFORMABLE OBJECTS

The simulation stage is also important in determining the discrepancy between a simulated scenario and its real world counterpart, thus the subjective plausibility of a physically based animation. Comparing to the rendering stage, the simulation stage may play a more important role by determining the overall movement and any possible deformation of objects. The temporal change of the exterior shape of objects provide much information that people can perceive. Therefore, understanding which and how factors in the simulation stage can influence the discrepancy and thus people's perception is useful.

In Study III, we study factors that may influence the perception of physically simulated deformable objects. Due to the complexity of physically based dynamics models of deformable objects and the variety of perspectives of people perceiving a deformation, there are many potential important factors in the simulation stage that can influence the visual plausibility and thus the perception of an animation with deformable objects. Therefore, any improvement in such study can be helpful for artists and researchers in their future work.

4.1 Motivation

In this study, we investigate the resolution of meshes used in simulation as a factor in influencing the perception.

Dynamics models used in the simulation of deformable objects could be the most determinant factor in influencing the visual plausibility of an animation with soft bodies. However, the dynamics model is not the major factor we study here. There are mainly two reasons. First, researchers have compared from mechanical point of view, various different dynamics models, not only among physically based ones, but also with other heuristic models [44]. Second, these models each has its own advantage and disadvantages, for example, some preserve the volume of mesh model while others can handle inverted elements. Usually, specific models are chosen in an application for certain reasons. Users may have options among several models, but they may not be able to or do not want to manipulate the model themselves.

Similarly, we do not study the parameters which can influence the dynamical properties in various dynamics models, for example, the Poisson's ratio. There has been comprehensive studies on the measurement of these parameters for materials in the real world.

In physically based simulation, the mesh used in dynamics can be different from that used in rendering. Usually the mesh used in simulation has less polygons and thus less DOF. We call it ***simulation mesh***. In contrast, the mesh used in rendering has more polygons and thus more details. We call it ***rendering mesh***. Once the vertices on the simulation mesh is updated. The position of vertices on the rendering mesh can be interpolated. This decoupling of simulation and rendering meshes can greatly reduce the computational cost while preserving the details in rendered animation.

Simulation meshes for the same object usually bound the spatial domain of the object. So they have about the same shape and volume. The difference is how many vertices and polygons are used to specify the boundary of the volume. We call the ones with more vertices, thus DOF and details, ***high resolution*** meshes. In contrast, the simulation mesh with less vertices is of ***low resolution***.

Figure 4.1 compares two animations with the same rendering mesh but simulated using meshes of different resolutions. Obviously, the resolution of the simulation mesh determines the physical accuracy and the computational cost of the dynamical simulation. Comparing to rigid body simulation, the resolution of mesh used in simulation is thus more important.

In Study III, we focus on the effect of resolution of simulation mesh on perception of

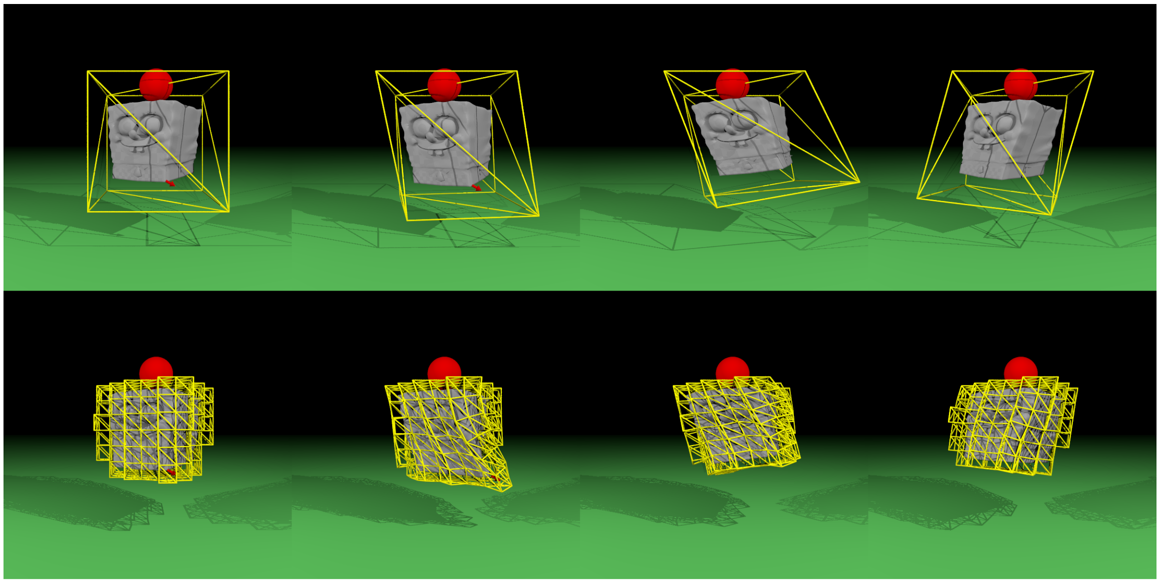


Figure 4.1: Frames taken from two animations of the same rendering mesh of Spongybob simulated using meshes of low resolution(top) and high resolution(below). Yellow lattice are the exterior edges of the simulation meshes.

the simulation of deformation. The main reasons as discussed above are: first, the mesh resolution may have direct influence on the visual details in the final animation; second, comparing to dynamics models and other factors, the mesh resolution is a more practical factor to control due to its easiness to manipulate and that it can be changed continuously.

4.2 Study III Experiment Design

4.2.1 Factors

Intuitively, the higher resolution a simulation mesh has, the more details it can provide. Correspondingly, it consumes more computation. If we start off by using a high resolution mesh and gradually reduce the mesh resolution to generate different animations, people will eventually notice the changes in the rendering mesh in final animation caused by such reduction. A common question is how much we can reduce the mesh resolution without introducing too much error in the final animation that can be noticed by the viewer. Thus we study the mesh resolution as a factor, reducing which may cause people to notice the difference between an animation and the referencing one that uses the mesh of the highest resolution.

Obviously, the resolution of a mesh is high or low when it is compared against another mesh. Thus the mesh resolution as a factor only has meaning in a relative sense. Another factor which is also relative and may interfere with the effect of the mesh resolution is the relative distance of a mesh to the viewer, or the size of an object to the human eye in terms of opening angle. For example, when the object is very far from the viewer that its size is so small for people to discern any details. The effect of the mesh resolution can barely be noticed either. In our current study, we eliminate such influence by using models of the same size and at the same distance to the viewer.

In the classic FEM models for deformation simulation, both tetrahedral elements and hexahedral elements are used. In computer graphics, tetrahedral mesh is a natural choice because it is more compatible with triangle surface mesh. It is also more straight forward to generate tetrahedral volumetric mesh from a triangle surface mesh whether preserving the surface geometry is necessary or not. Another benefit of using the tetrahedral mesh is the easiness to model the surface of the volumetric mesh such that it follows the shape of rendering mesh closely.

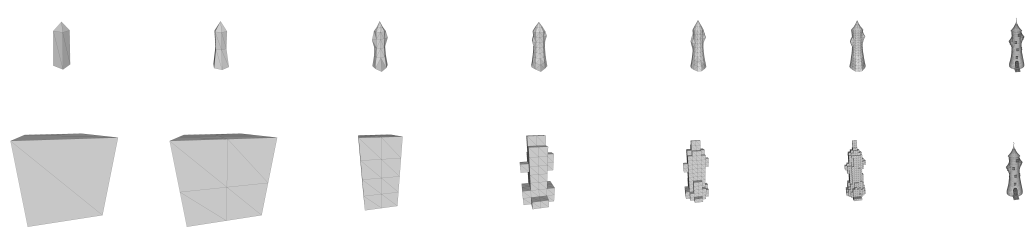


Figure 4.2: Tetrahedral mesh and hexahedral mesh in six mesh resolutions of the same rendering mesh.

Figure 4.2 shows volumetric meshes of different resolutions using tetrahedra and hexahedra. To generate these tetrahedral meshes, we start from a low-resolution approximation of the rendering mesh or the rendering mesh itself and take advantage of different modeling skills like decimation of vertices, re-meshing, subdivision, edge collapse, etc. We adjust the size of triangles on the surface such that they are at different levels for different mesh resolution. We then generate the volumetric model using Tetgen and preserve the surface geometry. We name the six mesh resolution levels 1 to 6 for now.

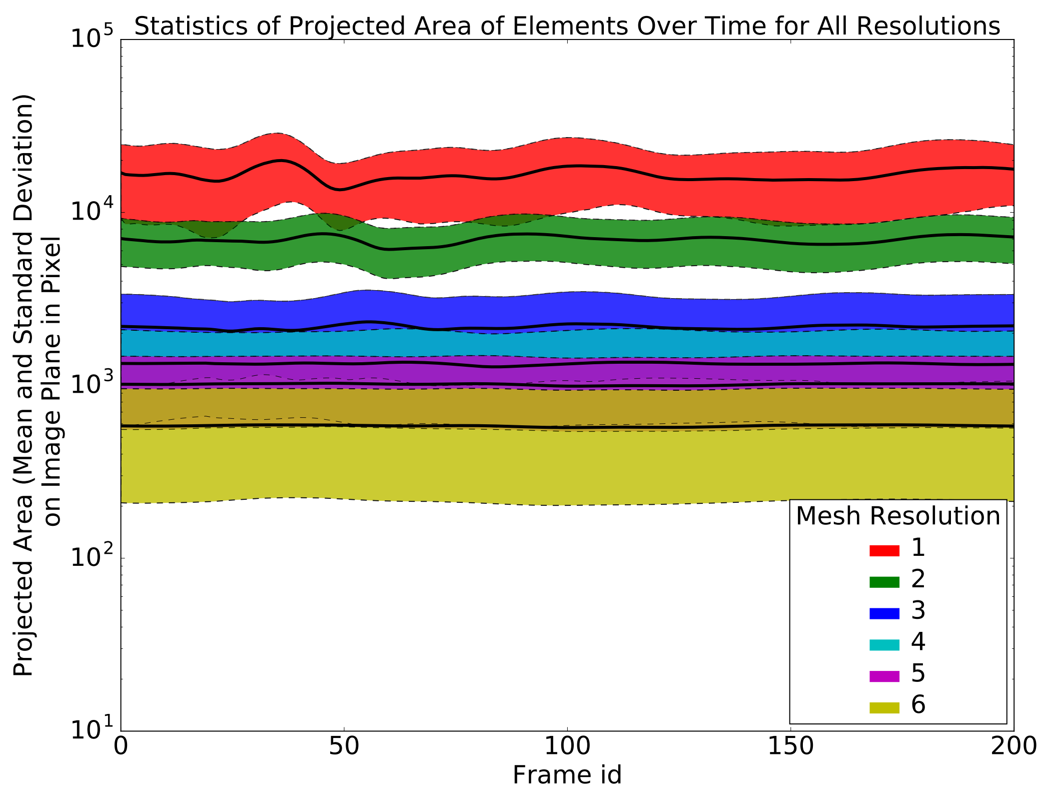


Figure 4.3: Statistics of projected area of tetrahedra in tower model of different resolution over time. Thick lines represent the mean of projected area over time. The dashed lines above and below each thick line and the shaded area between them represent the range of mean $\pm$ standard deviation. (Mesh resolution id just for annotation)

Apparently, such modeling task takes extra labor. Another problem with the tetrahedral model generated this way is that the elements may be of bad shape (e.g. very large or small aspect ratio). To numerically measure the goodness of such modeling method, we project each tetrahedral element on to the image plane. We statistically measure the area of the projection of all elements. During the simulation, all elements can deform and the statistics can change. Figure 4.3 shows the statistics of projected area for meshes of different resolution under the simulation of the same scenario in which the tip of the tower is pulled aside in the beginning of the simulation.

We firstly notice in Figure 4.3 is that the shaded area for mesh of resolution 1 and 2, and that for mesh of resolution 3, 4, 5 and 6 intersect with each other. This means that although the average projected area of elements in meshes of different resolution are apart from each other, some elements in different meshes are still of relatively close projected area and even volume. Meanwhile, the shaded band can be very wide for most resolution. This stands for the fact that even in the same mesh, the size of tetrahedra can vary in a wide range. It is also easy to see that the thick lines which represent the average projected area are not evenly distributed although they are in the correct order as expected. The last thing we notice is that the thick lines for some resolution can be squiggling over time. This means that during the simulation, the overall projected area of elements in a mesh change drastically. This is mainly because many elements in the mesh are of poor shape. Thus a stress in a certain direction may force the shape of elements to deform a lot or lose volume. These disadvantages all prove that the modeling method mentioned previously is not a good candidate to design simulation meshes of clearly different resolution. An experiment on mesh resolution based on models designed using this method can be not generalized well.

To generate simulation meshes of clearly different and evenly distributed resolution, we decide to use hexahedral elements. To generate simulation mesh of a certain resolution, we first choose a size for the hexahedral elements. We then decompose the bounding cube of the rendering mesh into elements of the chosen size at regular positions. We keep those elements that intersect with the volume of the rendering mesh and remove those outside the shape. The exterior surface of the simulation mesh do not follow the shape of rendering mesh as closely as the tetrahedral meshes. But the DOF can still be controlled as required and the collision can be handled using the rendering mesh without a problem. Another benefit is that the entire modeling process can be automated because there is no need for adjusting the simulation models in detail to match the rendering mesh.

In this study, the element size is chosen to be $BoundingCubeDimension/2^n$. And we define the ***mesh resolution*** in our context to be the denominator 2^n in the previous formula. For example, the lowest resolution is 1. Element in a mesh of resolution 1 is the entire bounding cube. For mesh of higher resolution, we regularly subdivide the elements in all three dimensions.

We create another six meshes of different mesh resolution as shown in Figure 4.2. Again, we measure the statistics of the projected area of all elements in each mesh during the same simulation used to generate Figure 4.3. The result is shown in Figure 4.4. The improvement is obvious. All six bands are relatively flat and evenly distributed.

A free-flying deformable object can have both the rigid displacement and deforming components. The translational and rotational components in the rigid displacement make the observation task more difficult. In this study, we want to focus on the effect of mesh resolution on the pure deforming components of an object. So we fix some vertices of each model to rule out the 6 DOF corresponding the rigid motion. According to our previous studies, the texture used for rendering can also influence people's perception. So we use the same low-contrast checkerboard pattern to render all models. In our following analysis, we will discuss other potential factors we discover during the study which may influence the effect of the mesh resolution.

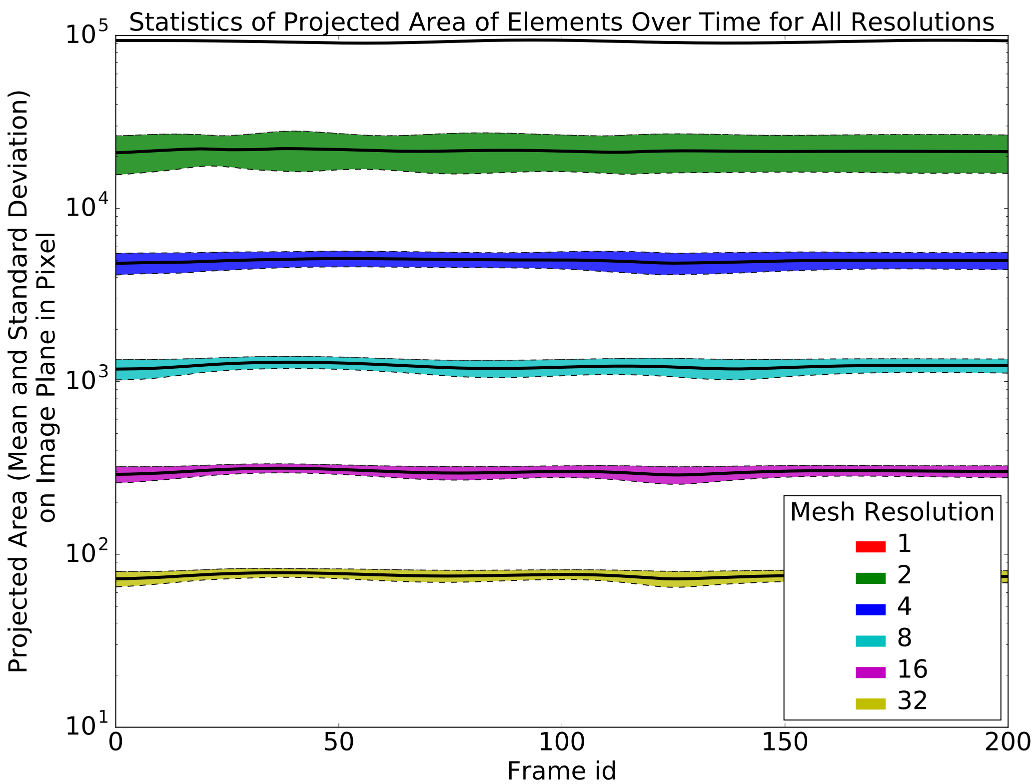


Figure 4.4: Statistics of projected area of hexahedra in tower model of different resolution over time. Thick lines represent the mean of projected area over time. The dashed lines above and below each thick line and the shaded area between them represent the range of mean $\pm$ standard deviation. Notice the deviation for mesh resolution 1 is zero.

4.2.2 Stimuli Preparation

Similar to previous studies, we use Blender for modeling, VegaFEM and Bullet for dynamics engine, and POVRay for rendering. In the example stimuli that we display at the beginning of the experiment, we also visualize the exterior edges of the simulation mesh as a bounding cage, as shown in Figure 4.1. This is to help people to understand the concept of mesh resolution. In formal stimuli, only the rendering mesh is visualized. We also render a red sphere to cover the fixed vertices and to clearly show viewer which area is being constrained. Any external force in the simulation is visualized as a red arrow. Each stimulus is of two seconds long. Videos are of the resolution 1024×768. They are displayed with a 23 inch screen. We use Psychopy to display the stimuli and collect feedback.

4.3 Study III Experiments and Analysis

There are two questions we try to answer in this study. The first is how much we can reduce the mesh resolution from a high-resolution referencing mesh before people can clearly notice the difference resulted in the animation. Another question is whether the answer to the previous question is uniform in different scenario. In other words, can the simulation scenario influence the amount of reduction we can perform before people notice the difference in the animation?

4.3.1 The Design of Stimuli

We use the 2AFC protocol in this study. In such experiment scheme, each stimulus is paired with an animation of the same scenario(i.e. same object and same external force) but simulated using a mesh of the highest resolution. Participants are told to choose the animation that is rendered from a simulation using a mesh of the lower resolution or with less DOF. The two stimuli are displayed in a random order. All stimuli pairs are displayed in a random order. Simulations at 6 mesh resolution levels(1, 2, 4, 8, 16, 32) are used. The number of times the stimuli at these resolution levels displayed are 2,6,8,8,8,8. The correctness of participants rating at different mesh resolution levels can then be used to fit psychometric curve. The index JND can be used to represent how much reduction can be done without resulting an animation obviously different from that of the highest resolution.

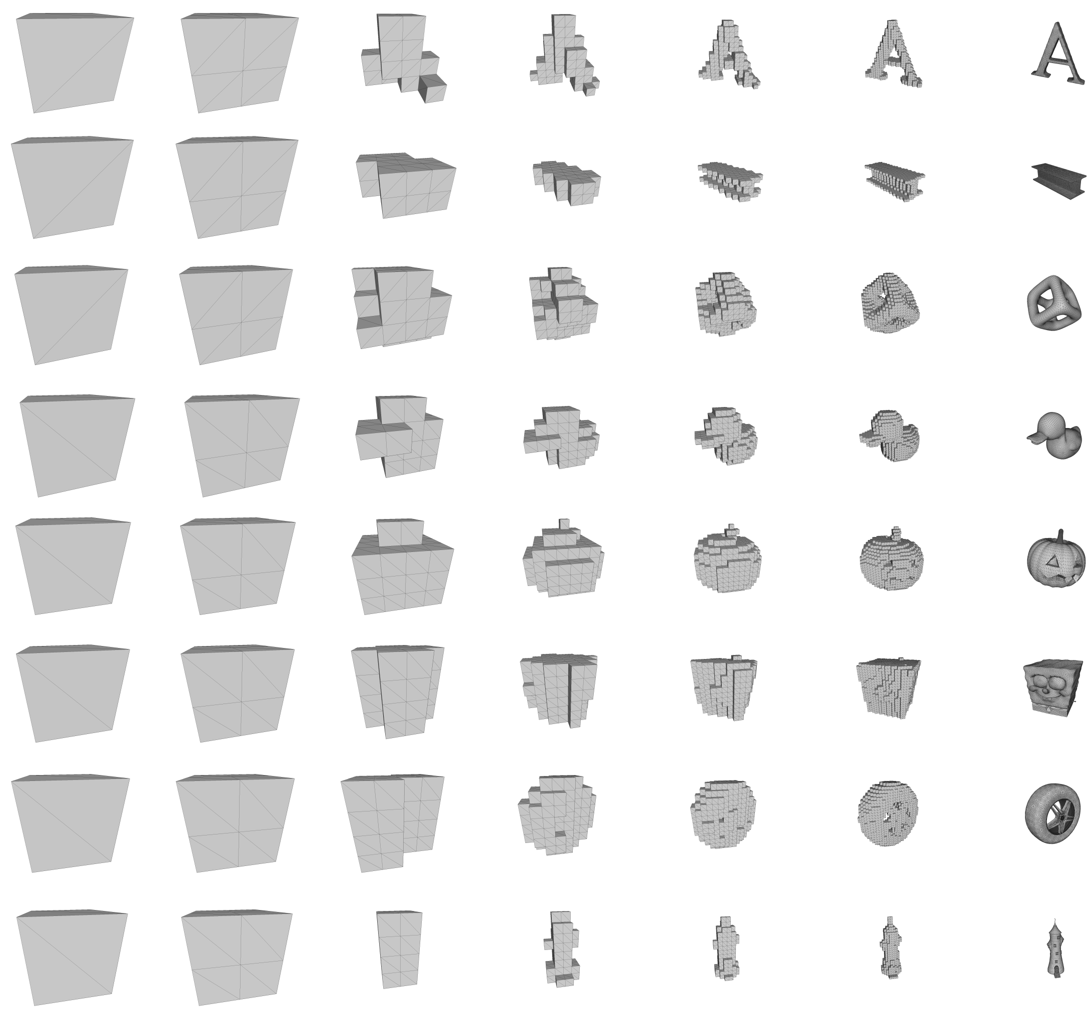


Figure 4.5: All eight shapes(rightmost column, from top to bottom: A, Beam, Cube, Duck, Pumpkin, Spongybob, Tire, Tower) and their hexahedral simulation meshes at six different levels(from left to right: 1, 2, 4, 8, 16, 32).

Figure 4.5 shows all eight different shapes and the corresponding hexahedral simulation meshes at six different levels. We use them to create two separate experiments.

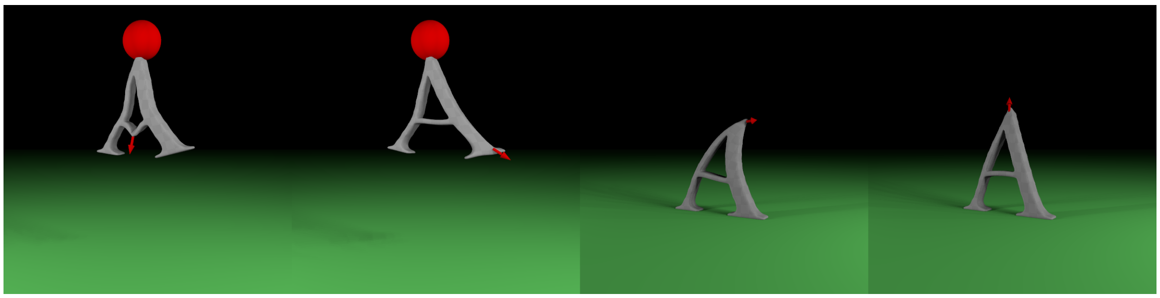


Figure 4.6: Frames taken from the animation of the four scenarios with A (from left to right: A-MiddleDown, A-RightLegRight, ABottom-TipRight, ABottom-TipUp).

In the first experiment, we only use the shape A. We simulate four different scenarios. In the first two scenarios, we fix the top of A and pull the middle bridge of A downward and pull the right leg of A to the right separately. In the other two scenarios, we fit the two legs of A on the ground. We then pull the tip of A to the right and upward separately. All external forces are applied only in the first 20 time steps of the simulation. Figure 4.6 shows frames taken from animation of these four scenarios.

In the second experiment, we use all eight shapes. We create one scenario for each shape by fixing part of the model and apply external forces on a different part of the shape in the beginning of the simulation for 20 time steps. Figure 4.7 shows frames taken from animation of the eight scenarios.

We explain to each participant the concepts used in our study. To help them to understand the concepts, they can also play with an interactive program in which they can move the vertices of the simulation meshes of different resolution and the rendering mesh will deform correspondingly. Before the formal experiment, example stimuli with bounding

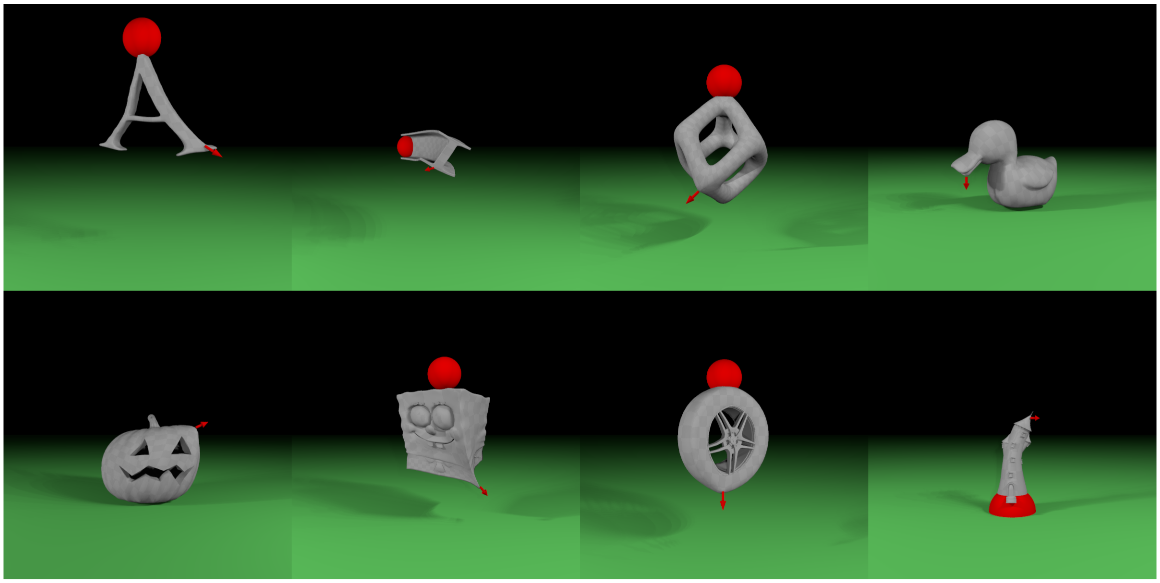


Figure 4.7: Frames taken from the animation of the eight scenarios with different shapes (top row, left to right: A, Beam, Cube, Duck; bottom row, left to right: Pumpkin, Spongybob, Tire, Tower).

cages are displayed to help them to get familiar with the scenario in other stimuli. Then, the participants begin watching stimuli and giving feedback.

4.3.2 Results and Analysis

We recruit nine people to participate the first experiment. Their ages range from 19 to 28. All participants are naive to the study. Most participants are not familiar with the concepts used in our study before we explain to them in the experiment.

Given people's answer for each pair of stimuli, we can compute the correctness of people's answer at each mesh resolution level for all scenarios. We further use Psignifit [16] to fit a logistic psychometric curve for each scenario which represents the correctness of people's answer as a function of the mesh resolution. This correctness function has a value close to 50% for highest mesh resolution 32, which is also used as referencing stimulus. This is the correctness of guessing. As the mesh resolution is reduced, the simulation result becomes more and more different from that of the highest resolution. Thus, it is easier for people to tell the difference between an animation rendered from a simulation using low resolution mesh and another one using a high resolution mesh. Assuming no other interfering factors, it is also easier for people to make a correct choice based on the observable difference.

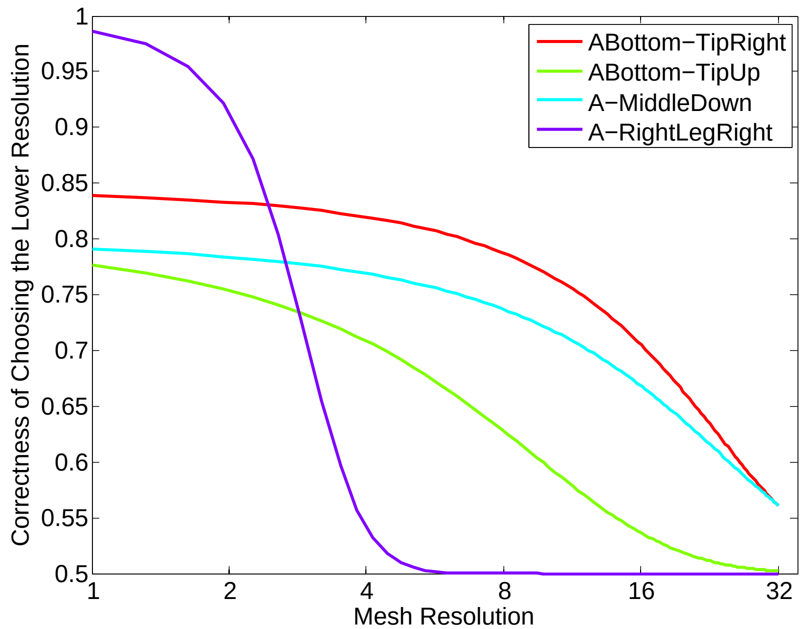


Figure 4.8: Psychometric curves of the four scenarios in the first experiment.

Figure 4.8 shows the psychometric curves for the four scenarios in the first experiment using data merged from all participants. The first observation of the figure is that all curves go down as the mesh resolution increases. This is consistent with the intuitive assumption that ***reducing the mesh resolution can increase the possibility for people to tell animation that uses the low resolution mesh from the referencing one that uses the highest resolution mesh***. However, three of the four curves are not close to the shape of a logistic function. After investigating the raw data, we believe this is mainly caused by people's misunderstanding of the concept mesh resolution and the imperfection of several specific stimuli. Due to the discrepancy of simulation using meshes of different resolution, the objects in some animation using low resolution meshes may have larger or faster deformation than the animation using high resolution meshes. People who do not fully understand how the mesh resolution plays a role in the pipeline may mistakenly recognize the larger deformation as a symbol of high resolution meshes being used.

Even though some meshes are not in perfect shape, we can still tell that their horizontal location are not close to each other. This hints that the reduction of mesh resolution in the four scenarios may have different effect. In other words, the different ways of disturbance in this experiment may be a factor that can significantly influence the effect of mesh resolution. To numerically analyze the influence from the potential factor, we again use the Just Noticed Difference(JND) threshold like in previous studies. The JND threshold in this experiment setting is the mesh resolution level that corresponds to the 75% correctness on the psychometric curve. We can approximately see that the JND of the four curves in Figure 4.8 are apart from each other. We further fit the psychometric curves and JND for all participants and all scenarios. We perform a one way repeated measure ANOVA on the JND thresholds of all participants. The different disturbance in the four scenarios is found to have a significant effect on people's JND $(F_{3,24} = 9.349, p < 0.0003)$.

A post hoc analysis shows that the JNDs for the A-RightLegRight and ABottom-TipUp scenarios correspond to smaller mesh resolution. This means that comparing to the other two scenarios, the mesh resolution can be reduced to a much lower level before people can notice the effect of such reduction with 75% correctness. By analyzing the simulation of these two scenario using mesh of highest resolution, we find there are mainly simple deformation in them. For example, the ABottom-TipUp only contains vertical scaling because the way the tip is pulled up. In the scenario A-RightLegRight, after being pulled, the model has a rotation-like overall deformation which does not contain much high-curvature local deformation. The deformation like these can still be reproduced by using simulation mesh with a low resolution. In contrast, there is local or complex deformation in the other two scenarios. In the A-MiddleDown scenario, the thin middle bridge deforms locally and drastically after being pulled downward. In the ABottom-TipRight scenario, the model bends and has a very curvy leg after its tip being pulled to the right because both its legs are fixed on the ground. These detailed deformation can not be easily reproduced by using low-resolution meshes. This observation supports the hypothesis that ***the reduction of mesh resolution is easier to be recognized when the deformation simulated using mesh of high resolution is complex or has many local details***.

In the second experiment, we recruit another 7 people. Their ages range from 19 to 32. They are all naive to the study. We again fit the psychometric curves using the merged data as well as for each individual. It needs to be mentioned that the feedback data of one participant is determined to be an outlier and is not used in fitting the curves in Figure 4.9. The data and our communication after the experiment indicates that the participant does not understand the concepts correctly. His data shows randomness and it can barely be used to fit psychometric curves.

Nonetheless, we can still see clear pattern in Figure 4.9 plotted using the data of the other six participants. Although the curves of two scenarios, Tire and Tower, are not in the shape of a logistic function, the other six curves are all in a good shape. Also, all eight curves are monotonically decreasing as mesh resolution increases. This is consistent with our observation in the first experiment. Further examination of the stimuli of the Tire and Tower scenarios and the participants feedback on the two scenarios point to similar culprits as in the first experiment. The stimuli at the mesh resolution levels 2 and 4 in both scenarios show larger, faster or very different deformation comparing to the stimuli at the highest mesh resolution level. This indicates potential improvements in our future work.

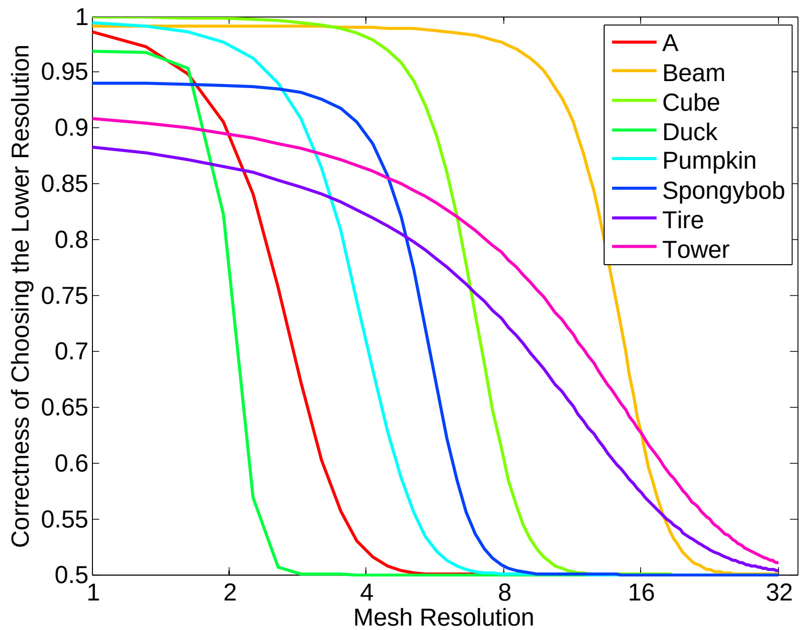


Figure 4.9: Psychometric curves of the eight scenarios in the second experiment.

We again observe the interesting pattern that the horizontal locations of the curves are separated from each other. By fitting the JNDs for all participants and scenarios and performing the one way repeated measure ANOVA, we find that different shape and disturbance is found to have a significant effect on people's JND $(F_{7,35} = 3.285, p < 0.0086)$. A post hoc analysis shows that the JNDs of curves on the two sides of Figure 4.9 are significantly different from each other. For example, Beam is significantly different from Duck$(p < 0.0098)$. Duck is significantly different from Cube$(p < 0.059)$. Although the current experiment can not provide enough data to claim that the curves in between are significantly different from each other, we can still see an ordering of them, which hints for a potential pattern. By replaying and comparing the stimuli of the eight scenarios, we find a pattern similar to that observed in the previous experiment. The deformation in the Duck scenario is mainly the bending of the head of the duck due to its beak being pulled downward. There is not much local curvy deformation of high complexity. As a result, such deformation can be reproduced even with mesh resolution 2, with the head and beak almost in one element and the body in other elements. This is consistent with the fitted curve. In contrast, it is obvious even in the single frame in Figure 4.7 that the edge of the beam squiggles after one end of the beam is pulled aside. The edge becomes more wavy in the following frames in the animation. The cube is different from the A in that the motion of model A is mostly in a plane, where the cube is like a 3D cage. Although the cube can also rotate after being pulled, it has more wavy propagation of deformation energy than A. This observation again supports the hypothesis mentioned previously that the reduction of mesh resolution of simulation that is more complex or has more details can be recognized easier.

4.4 Study III Conclusion

In this study, we perform two experiments to investigate the effect of resolution of simulation mesh on people's perception of animation with physically based deformation. We further study how such effect can be influenced in different scenarios.

We first prove that reducing the resolution of simulation mesh can cause people to notice the difference in the animation due to the change of simulation mesh. The more reduction is applied, the easier it is for people to notice such difference. We further observe that the effect of mesh resolution reduction can vary in different scenarios. Our data supports the hypothesis that the more curvy details a deformation using high resolution mesh has, the easier it is for people to notice the difference in animation caused by reducing the mesh resolution.

Our study proves the effectiveness of the model reduction in improving the simulation efficiency without harming the visual plausibility. It can be seen from several scenarios in our study that the correctness of people recognizing the animation with the true low resolution does not increase to more than 50% until the mesh resolution is reduced to 8 or even 4 from 32. Our study further provides helpful insight of reducing mesh resolution. For example, an analysis of the ideal scenario being simulated to determine the observable deformation complexity is necessary before any model reduction can be applied. Depending on the complexity of deformation or the amount of curvy details of deformation, one can have an approximate idea of whether aggressive mesh resolution reduction is noticeable by people. All these observations can help artists in choosing a better simulation mesh to balance the computational cost and visual details.

Our study so far is only a primary step in investigating the factors in the simulation stage of animation with deformation. To further verify the hypothesis proposed in previous section, we need to recruit more participants and design more specific scenarios to have enough samplings and more sound evidence. Meanwhile, to facilitate the numerical analysis and allow the study result to be more useful, a metric that measures the visually deformation complexity or the amount of curvy details may be necessary. Such metric can be used as an objective index of deformation complexity. Some potential candidates are metrics using modal analysis, statistical analysis on the simulation data and mechanical analysis on the strain tensor. Of course, we need to improve the design of experiment and stimuli. We may consider recruit more participants like experienced artists or people who are familiar with the animation production pipeline. Also, we need to reduce the influence of other unexpected factors like narrowing the difference in the extent or speed of overall displacement in stimuli with different mesh resolutions. Finally, it is interesting to see how the effect of the mesh resolution can change in more complex scenarios like those with free-flying objects.

5. FACTORS IN THE SIMULATION OF ANIMATIONS WITH RIGID BODIES*

In this section, we focus on factors in the simulation of rigid bodies. More specifically, in Study IV, we study the effect of factors in the simulation of large scale rigid bodies. The simulation of a single rigid body is trivial and the standard rigid body dynamics model can be used. What interests people is the simulation of a large number of rigid objects, for example, the piling of lots of rigid objects.

5.1 Motivation

Large scale physically based simulation is widely used in entertainment. As computing power has increased, this usage has become even more common. In addition, a number of new techniques for physically-based simulation have also increased the applicability of these methods. Some of these methods have the advantage of improved speed, but at the cost of reduced accuracy.

Off-line simulation methods exist in almost all areas and can achieve very accurate simulation results, but they are often resource-intensive. Furthermore, such physical accuracy is not always necessary, particularly for entertainment applications. Visual plausibility, which may deviate from physical plausibility, is the actual standard that is eventually used. Intuitively, it is often very difficult for a human to tell whether a moving object or deforming scene under simulation is physically accurate or not. People may accept that a simulation result is accurate if the distortion is small. In large scale physically based simulation, attention is also distracted by multiple objects. Thus it may be more difficult for viewers to tell if distortion exists. If these two hypotheses, that people don't notice small deviations and that with increasing numbers of objects people accept larger deviations, are

*Part of the data reported in this chapter is reprinted with permission from "Believability in simplifications of large scale physically based simulation" by D. Han, S.-w. Hsu, A. McNamara, and J. Keyser, 2013. *Proceedings of the ACM Symposium on Applied Perception*, pages 99-106, Copyright 2013 by ACM.

true then physically inaccurate simulation can still be visually plausible. To achieve higher efficiency or more controllability, many algorithms assume intuitive hypotheses like these to be true in simulated scenarios. However, there is only a small amount of prior work either verifying the authenticity of such hypotheses or providing a metric to measure visual plausibility objectively.

O'Sullivan et al. [46] verify the authenticity of the hypothesis that people may not be able to detect small distortion in physically based simulation. They investigate subjective tolerance on several kinds of distortion in physically based simulation. Their work focuses on simple scenarios where only a few rigid objects collide with each other. In this case, viewers can concentrate on the colliding event and locate poor simulation. With hundreds of thousands of objects simulated at the same time, one would expect that a viewer's attention will be distracted by different events happening simultaneously. As a result, a viewer's tolerance threshold for distortion, if there is one, may be different or even unstable. It's possible that, while viewing large scale simulation, a viewer's judgment of visual plausibility may depend on a global moving pattern of objects instead of local physical plausibility. On the other hand, distortions mentioned in [46] are not always what an approximated simulation method would generate. For example, objects in or under a pile of objects are partially occluded by other objects and are seemingly under greater pressure, thus it may be easy for a viewer to accept the assumption that these objects do not transform during simulation. Thus "freezing" them would be a good approximation. In large scale simulation, more global statistical features may reflect the overall distortion level better. So, the verification result in [46] may not hold for similar hypotheses in large scale rigid body simulation.

As a first step, we only focus on large scale rigid body simulation. We firstly verify the authenticity of the following two hypotheses through a psychophysical experiment.

Hypothesis I: In large scale rigid body simulation, viewers may not be able to perceive distortion, such as that incurred by an approximated simulation method.

Hypothesis II: Assuming objects in or under a pile of objects to be fixed without transformation does not affect the visual plausibility of simulation.

5.2 Study IV Experiment Design

5.2.1 Factors

The major factor under investigation in Study IV is the rigid body dynamics models used in large scale rigid body simulation. As mentioned earlier, different models are based on different assumptions which are equivalent to the hypothesis to be verified in this study. Each hypothesis stands for a certain form of effect of this major factor. By verifying them, we can better understand the effect of the major factor on people's perception of animation with a large scale of rigid bodies.

In the experiments, we also investigate several other factors which may potentially influence the effect of the main factor: number of collisions under simulation, homogeneity of colliding object pairs, distance from scene to camera.

5.2.2 Rigid Body Dynamics Models Used in The Experiment

Similar to the overall appearance, the rigid body dynamics model is a categorical factor too. Different dynamics models can be seen as discrete sampling levels of the entire population of this factor.

We create animation stimuli using several approximated rigid body dynamics models as well as a classic impulse-based simulation method. We summarize them as *M1* to *M5* in this section.

***M1*:** As a benchmark, we generate stimuli using a classic simulation method [20] for all scenarios which will be discussed in the following sections. The classic simulation method iterates among four stages: collision detection, collision determination, collision response, and integration. Stimuli generated using this method are expected to have the highest visual plausibility and highest physical accuracy. But, this method is the most computationally expensive one. We label this method as *M1*.

***M2*:** Comparing to the classic simulation approach, approximated simulation methods consume fewer computing resources by simplifying some of the four stages or using different simulation schemes. Hsu and Keyser replace the classic collision detection and collision response computation with a table lookup-like scheme [27]. This method can save up to 99% of the time to simulate one step. Their high performance is at the cost of significant pre-computation. In pre-computation, statistical data are collected by performing a series of trials. The statistical data is a compact representation of an object's response to collision events. In real time simulation, collision responses generated with this simplified method are statistically close to actual responses when a lot of objects exist. However, with regard to any single object, there is high likelihood for distortion of the results. This method is used to verify *Hypothesis I*. We label this method as *M2*.

***M3*:** Rather than uniformly modifying stages in simulating each object, Hsu and Keyser improve the overall performance by stopping simulation of objects that satisfy certain conditions [28]. Speed-ups achieved in this paper range from 1.2 to 6.0 times depending on specific scenarios and various other factors. Distortion in such simulation comes from the "frozen" objects. A large pile of objects with some of them having such distortion may still be plausible. This method is used to verify *Hypothesis II*. This method (see the original paper for details) allows one parameter to be set to trade off between simulation speed and accuracy. We use a value of 0.5 in our examples. This method is labeled as *M3*.

***M4*:** Furthermore, an LOD-based method that combines statistical and classic methods is also used to simulate some scenarios. We label this method as *M4*. According to an object's visual importance or necessity in the scene, *M4* chooses between *M1* and *M2*. The distortion is suppressed accordingly. In our implementation, *M4* uses *M1* if viewing distance of the simulated object is less than some threshold (we use 140 units, where the size of an object is approximately 15 units in diameter). Otherwise, *M2* is used.

***M5*:** Finally, we also generate animation stimuli using *M5* where the collision model of each object is simplified to be a single sphere.

5.2.3 Measurements

To verify authenticity of *Hypothesis I & II*, we design a series of scenarios each of which is simulated with different simulation methods. Subjective visual plausibility evaluations of each scenario is then collected from viewers. Statistical analysis is performed to verify the truthfulness of *Hypothesis I & II*. To collect subjective visual plausibility, as in [46], we ask volunteers to watch each stimulus and afterward rate the realism of the physical behavior seen in the stimulus using a 7-point Likert scale ranging from 1 (totally unrealistic) to 7 (completely realistic).

Our goal is to verify the truthfulness of these hypotheses and to investigate factors which may affect such results. We leave the measurement of a tolerance threshold of a specific metric of each factor to future work. As a result, we did not apply the randomly interleaved staircase design [10]. Instead, all our animation stimuli are generated ahead of the experiment.

In our experiments, we also use an eye-tracking device to record viewers' gaze position while they are watching the script. The device we use is faceLAB$^{\mathrm{TM}}$5 from Seeingmachines. The screen we use to play the script has size 52cm$\times$32.5cm with resolution 1920$\times$1200. The tracking device comes with two IR cameras mounted on the table. To ensure the tracing accuracy, each viewer performs a calibration prior to beginning the script. The calibration takes less than 10 minutes and no further calibration is performed during the script. The gaze position data as well as the rating data are recorded and used to measure the visual plausibility of each of these stimuli.

5.2.4 Stimuli Preparation

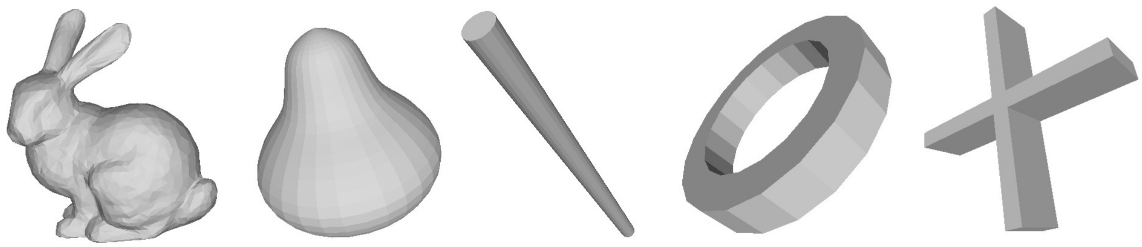


Figure 5.1: Models we used in scenario simulation

Model Name	Bunny	Roly-Poly	Rod	Ring	Cross
# of faces	3782	2560	240	160	44
# of vertices	1893	1312	122	80	24

Table 5.1: Geometric details of model used

We generate 32 different animation stimuli in total. They are simulated in 8 scenarios. These stimuli are further classified into three groups with similar scenarios in each group. Each stimulus lasts for less than 20 seconds. Collisions between object-object and object-environment pairs are simulated. Figure 5.1 shows the models used. Table 5.1 provides a description of each model used. Figure 5.2 shows a screen shot from one of the animation stimuli. Tab.5.2 lists the seven scenarios in detail.

We create a single script which consists of opening, introduction, displaying all stimuli, rating interface after each stimulus, and ending. This script is displayed to each viewer. PsychtoolBox [7] and Matlab was used to create this script. This toolbox allows easy manipulation of stimuli and provides an easy way to add an interactive script and to collect

Group ID	Scenario ID	Scenario Description	Simulation method used				
			classic	statistic	freeze	LOD	single-sphere
I	1	Falling rings collide with pole	$\surd$			$\surd$	$\surd$
	2	Bunnies collide with Roly-poly	$\surd$			$\surd$	
	3	Bunnies collide with rotatable crosses	$\surd$			$\surd$	
II	4	Bunnies fall on ground	$\surd$	$\surd$			$\surd$
	5	Shoot Bunnies	$\surd$	$\surd$			
III	6	Rings pile in box	$\surd$		$\surd$		$\surd$
	7	Bunnies pile on ground	$\surd$		$\surd$		

Table 5.2: List of scenarios generated to verify *Hypotheses I and II*.

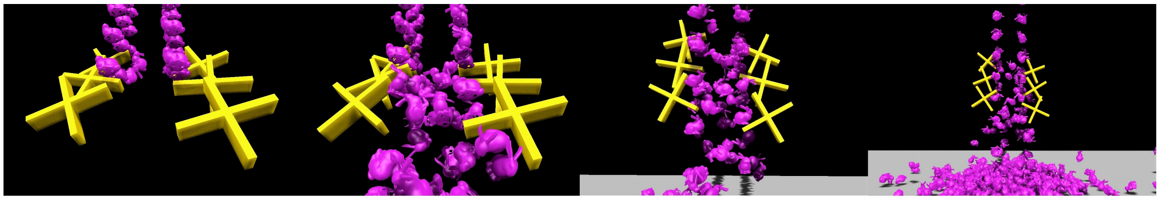


Figure 5.2: Screen shots of one stimulus used in our experiment. The four screen shots from left to right are sampled from one stimulus of the bunny-cross scenario. In this scenario, bunnies fall and collide with crosses which are only rotatable. A very large number of collisions happen during the simulation process.

feedback. In the script, a textual introduction is first displayed to the viewer. Three example stimuli are then displayed to help the viewer become familiar with the viewing-rating scheme. The actual experiment stimuli come after examples. Each stimulus is displayed by itself in the center of the screen. After each stimulus is displayed, a textual screen prompts the user for a Likert rating. Viewers then input their rating to proceed to the next stimulus. Viewers are also asked to input ratings for the examples to help them to become comfortable with the Likert scale rating process, but these results are not used.

Animation stimuli from the same group are displayed together, but the order of stimuli in the group is randomized for each viewer. Also, the order of groups is randomized. stimuli in some groups are played twice to measure the reliability of user rating. The order of repeated groups and stimuli in those groups are randomized, too. To restrict the total length of the script, not all groups are repeated.

Each stimulus is played individually, thus there is no side-by-side comparison. This is because our experiment focuses on measuring the subjective experience of a viewer when watching an animation. Side-by-side comparison of animations, however, aims to expose the difference or defects of one or both animations under comparison. While watching a movie, playing a computer game, or interacting in a virtual world, viewers are seldom given a benchmark for comparison. Instead, viewers have a subjective feeling according to their knowledge of the real world which the animation is similar to. Our purpose is to measure such subjective feelings of viewers.

5.3 Study IV Experiments and Analysis

5.3.1 Experiment IV.a : The Effect of Rigid Body Dynamics Models

5.3.1.1 The Design of Scenario

In this part of the experiment, we focus on verifying *Hypotheses I & II* under a common large-scale rigid body simulation environment. A typical setting for such simulations has hundreds to thousands of rigid bodies in the scene. Simulation parameters like gravity and friction are set in line with what might be seen in the natural world. The first two groups of scenarios are used to verify the first hypothesis and the third group is used to verify the second hypothesis. In group I scenarios, multiple types of objects exist and *M4* is used. *M2* is used to generate group II scenarios where only a single type of objects exists. *Hypothesis II* is verified in the group III scenarios, where piling situations are simulated.

Assuming *Hypothesis I* is true, even if simplification of some collisions may cause distortion, it can be assumed that such distortions do not affect visual plausibility. *M2* and *M4* make approximations to collision detection and response following this assumption.

Both *M1* and *M4* are used to generate stimuli for the three scenarios in group I. In the first scenario, more than 1000 rings fall from a certain height with randomized initial status and collide with 25 vertical and fixed poles on the ground. Both the swarm of rings and fixed poles are positioned densely. As a result, a large number of collisions happen when rings hit poles or the ground, and when rings pile on each other. Such collisions happen between rings, ring-pole pairs, and ring-ground pairs. Collisions occur throughout the entire simulation. We also simulate scenario 1 with *M5*.

The other two scenarios in this group are similar to the first one. In the second scenario, we place a roly-poly model in the shape of a Matryoshka doll on the ground. Then we shoot bunnies toward this model. Collisions happen between the roly-poly and bunnies, as well as among bunnies and between bunnies and ground. In the third scenario, we fix the center of six rotatable crosses. Hundreds of bunnies are then dropped from above and collide with these crosses and among themselves. Collisions also exist between bunnies and the ground. See Figure 5.2 for more screen shots. All three scenarios in the first group are designed to have large numbers of collisions among multiple types of objects.

In the two scenarios in group II, there is only one type of object. We use a bunny model in this case. *Hypothesis I* is assumed true and all objects are simulated using the same method. We simulate these two scenarios using both *M1* and *M2*. Furthermore, we simulate scenario 4 with *M5* as comparison. In scenario 4, we drop hundreds of bunnies from a certain height and let them fall freely to the ground. Bunnies at the same height are loosely positioned originally, so no obvious pile is formed when they hit the ground. But, there are still a large number of collisions among bunnies. Scenario 5 involves shooting bunnies toward the ground without colliding with any other objects before hitting the ground. Again, collision types here are limited to bunny-bunny and bunny-ground, but the number of collisions is still high.

Assuming *Hypothesis II* is true in simulation of large piles of rigid bodies, freezing objects inside a pile which may be under many other objects or occluded by others won't affect visual plausibility. According to this, *M3* improves the simulation efficiency and can also achieve an artistic effect.

Both scenarios in group III satisfy the condition mentioned above. In scenario 6, we place a rectangular box on the ground and toss rings from three directions into the box. More than 400 rings are simulated in this scenario. Rings pile quickly in the box and soon overflow from the box. In this scenario, many rings are occluded by the pile itself and the box, where it's natural to assume that *Hypothesis II* is true. Scenario 7 is similar to scenario 6. Instead of tossing rings, we drop a big array of bunny models to the ground.

Objects in this scenario are densely positioned so they are prone to form a pile. We use both *M1* and *M3* to simulate these scenarios. Again, *M5* is used to simulate scenario 6.

5.3.1.2 Results and Analysis

Thirty volunteers (12F/18M) participate in our study. Their age range from 20 to 33, with an average age of 25. Thirteen of the volunteers have a background in Computer Graphics, while the others are considered naive to computer generated animation. All viewers have normal or corrected-to-normal vision.

Before the experiment, each volunteer is told to rate each stimulus according to how realistic it is. They are also told that all objects simulated in stimuli are rigid bodies. They are told to focus on the physical behavior of objects only and ignore influence of other factors like color or lighting effects. Volunteers are explicitly told that some scenarios are animated with contrasting methods and some stimuli are displayed more than once. Almost all volunteers naive to graphics become comfortable with the viewing-rating scheme after rating the first three example stimuli. When the formal experiment begins, most viewers are observed to be able to concentrate on the stimuli and give a rating at once.

Analysis of Data in Group I

Figure 5.3 shows estimation of mean and 90% confidence interval of ratings for each stimulus over all viewers. In scenario 1 and 2, stimuli simulated with *M1* have higher average rating than those simulated with *M4*. In scenario 3, *M4* is actually a little higher than the classic simulation method *M1*. However, the difference between average ratings of stimuli is relatively small in the same scenario simulated with *M1* and *M4*.

A one-way repeated measures ANalysis Of VAriance (ANOVA) is performed to measure such difference. No effect of different simulation methods is found in scenario 1 and 3, see Tab.5.3, which means there is no significant difference between ratings for stimuli simulated with *M1* and *M4*. This result indicates that *Hypothesis I* can be assumed true

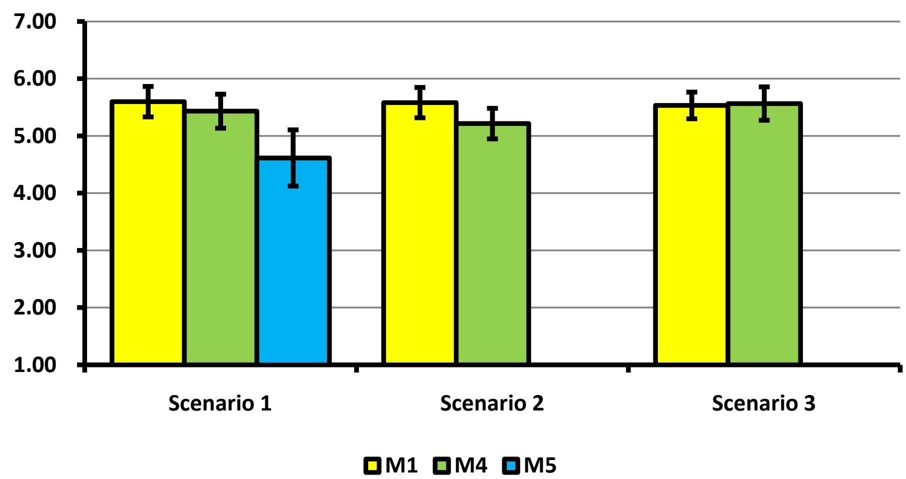


Figure 5.3: Estimation of mean and 90% confidence interval of rating for each stimulus in Group I over all viewers simulated with *M1* and *M4*

Scenario ID	F-value	Probability
Scenario 1	$F_{1,29} = 1.986301$	$p < 0.17$
Scenario 2	$F_{1,29} = 9.382353$	$p < 0.005$
Scenario 3	$F_{1,29} = 0.027645$	$p < 0.87$

Table 5.3: One-way repeated measures ANOVA F-value of Group I. The factor under test is different simulation methods. ANOVA is performed for each scenario separately. *M5* is not taken into consideration in this table.

under certain simulation environments, at least in scenario 1 and 3.

An interesting phenomenon is that the average ratings of *M1* and *M4* in scenario 2 has a difference larger than other scenarios. This indicates less confidence in assuming *Hypothesis I* to be true in this case. It also implies that factors affecting truthfulness of *Hypothesis I* may exist in this scenario. An investigation of stimuli shows that there are fewer simultaneous collisions in scenario 2 than in the other two. In scenario 2, bunnies are shot to a roly-poly. However, they bounce off in all directions. This decreases the number of collisions among bunnies. As a result, the number of objects or number of collisions in the scene may affect the truthfulness of *Hypothesis I*.

Different simulation methods can lead to different discrepancy patterns or different amounts of distortion. To further verify how *Hypothesis I* is affected by different simulation methods, we simulate scenario 1 with *M5*, a particularly poor substitute for simulation results. Again, estimation of mean and 90% confidence interval of rating for that stimulus is shown in Figure 5.3. As can be seen, the estimated mean of this rating is obviously different from that of stimuli simulated with the other two methods. A one-way repeated measures ANOVA is performed between *M1* and *M5*. The different simulation method is found to be a significant factor in the ratings($F_{1,29} = 17.2592, p < 0.00027$). The large distortion incurred by the over-simplified simulation method causes a rating much lower than the other two. In other words, *Hypothesis I* is not true if the simulation distortion is too high, even though it is true in the same scenario with other simulation methods. We will further evaluate this and the factor of number of collisions later.

Analysis of Data in Group II

Collisions in scenarios in this group are more homogeneous than those in group one because there is only one type of object in the scene. Figure 5.4 shows the estimated mean and 90% confidence interval of rating for scenario 4 and 5 stimuli simulated with *M1* and *M2*. Again, a one-way repeated measures ANOVA is performed. Results are listed in

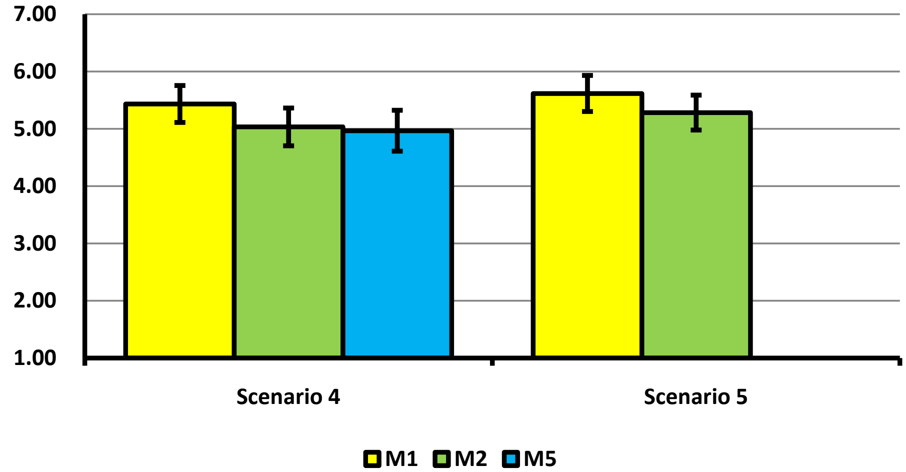


Figure 5.4: Estimation of mean and 90% confidence interval of rating for each stimulus in Group II over all viewers simulated with *M1* and *M2*

Tab.5.4.

From this data we find that the rating for stimuli simulated with *M1* and *M2* are not significantly different from each other. In other words, *Hypothesis I* can be assumed to be true in simple scenarios where only single types of objects exist. We shoot bunny models in both scenario 5 and scenario 2 in group I. The difference is that in scenario 2, bunnies collide with a roly-poly first and then fall on the ground. Whereas in scenario 5, bunnies are shot to the ground directly. Collisions in this case are not only more homogeneous but also more simultaneous. As a result, *Hypothesis I* can be assumed true with a higher significance.

It was somewhat surprising that the rating for *M5* in scenario 4 also has a high average (see *M5* in scenario 1). ANOVA analysis indicates no effect of different simulation methods between *M1* and *M5* in rating data ($F_{1,29} = 5.05694, p < 0.035$). This is to say that in such a simple scenario, *Hypothesis I* can be assumed true even for a simulation method with large distortions like *M5*. But in scenario 1, different simulation methods was found

Scenario ID	F-value	Probability
Scenario 4	$F_{1,29} = 2.351351$	$p < 0.14$
Scenario 5	$F_{1,29} = 3.258427$	$p < 0.10$

Table 5.4: One-way repeated measures ANOVA F-value of Group II. The factor under test is different simulation methods. ANOVA is performed for each scenario separately. *M5* is not taken into consideration in this table.

to be a significant factor. Thus, *Hypothesis I* does not hold for *M5* in scenarios in group I. One possible reason is that when collisions in scenarios are more homogeneous, it may be more difficult to people to judge whether they are all accurate or not. Thus, users may not be able to tell the distortion if they are all of the same type of distortion, even if the distortion is large. This implies using more simplified simulation methods rather than normal methods may be even more acceptable in applications with homogeneous collisions. We found these results to be somewhat surprising and counterintuitive.

Analysis of Data in Group III

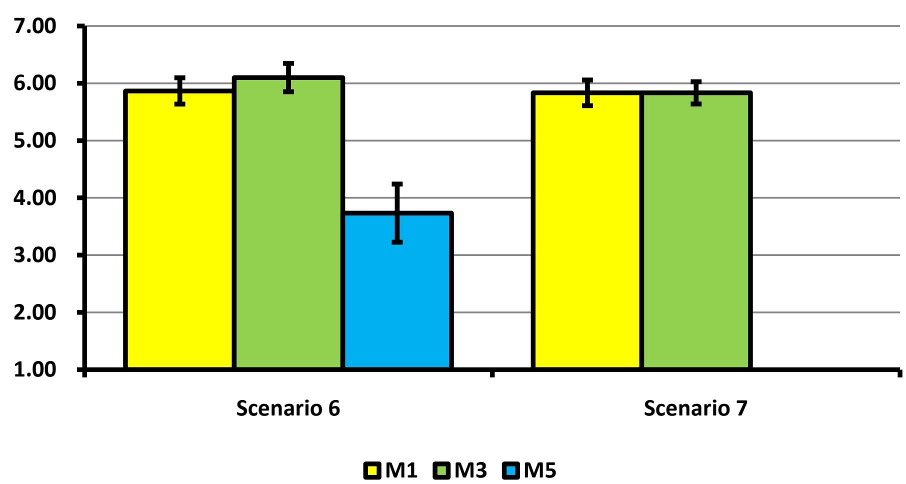


Figure 5.5: Estimation of mean and 90% confidence interval of rating for each stimulus in Group III over all viewers simulated with *M1* and *M3*

In this group of experiments, we verify the authenticity of *Hypothesis II* in two scenarios that involve piling of objects. Again, estimated mean and 90% confidence interval for each stimulus in both scenarios are shown in Figure 5.5. F-test values of one-way repeated measures ANOVA are given in Tab.5.5.

Scenario ID	F-value	Probability
Scenario 6	$F_{1,29} = 6.429864$	$p < 0.02$
Scenario 7	$F_{1,29} = 0.0001$	$p < 0.99$

Table 5.5: One-way repeated measures ANOVA F-value of Group III. The factor under test is different simulation methods. ANOVA is performed for each scenario separately. *M5* is not taken into consideration in this table.

Interestingly enough, *M3* has higher average rating than *M1* in scenario 6. They have about the same average rating in scenario 7. F-test values in ANOVA further prove such phenomena. Furthermore, in scenario 7, stimuli simulated with stacking simulation method *M3* has piles of objects with steeper slope than those in stimuli simulated with *M1*. However, the rating data indicates that people didn't find this difference significant. In other words, *Hypothesis II* can be assumed to be true in simulations of pile of objects as simple as in scenario 7.

To further verify the authenticity of *Hypothesis II* with different simulation methods, we simulate scenario 6 with *M5*. Results shown in Figure 5.5 clearly shows that *Hypothesis II* doesn't hold with *M5* which incurs too much distortion in this scenario. ANOVA between *M1* and *M5* ($F_{1,29} = 58.25601, p < 0.0000001$) also proves the falseness of *Hypothesis II* in this case. Unlike scenario 4, collisions are not homogeneous in this scenario. This again indicates that the factor of different simulation methods may affect the truthfulness of these hypotheses.

Eye Tracking Data Analysis

We also attempt to find an "objective" metric of visual plausibility from eye tracking data which is less affected by viewer's consciousness. Such a metric, if it exists, could be more helpful in measuring plausibility of simulation. To do this, we record volunteers' gaze position during the experiment. Viewers' gaze position on the screen is sampled 60 times per second. The sequence of gaze position data of each viewer is then filtered to compress noise. We compute the number of saccades, the average fixation duration, average gaze moving velocity, and average gaze moving acceleration of each viewer. We tested whether any of these factors might correlate with either the subjective ratings of viewers or the types of simulations.

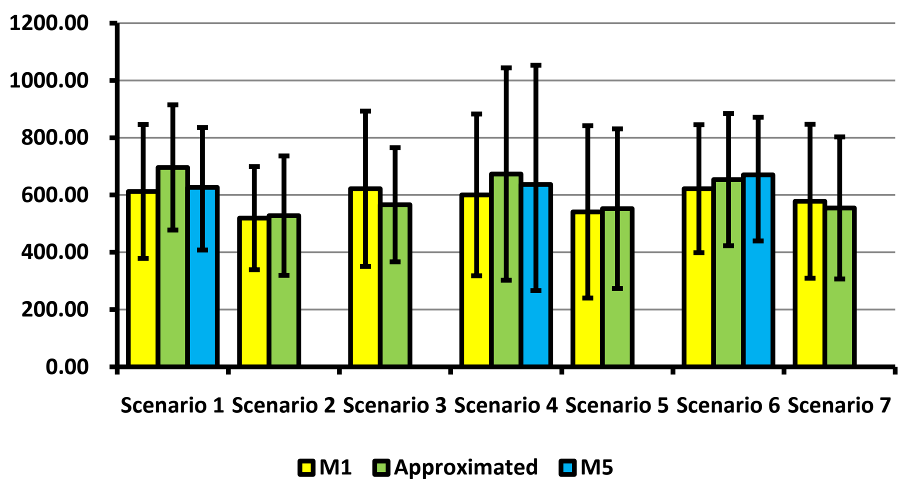


Figure 5.6: Average and standard deviation of average gaze moving velocity of viewers when watching different stimuli

Figure 5.6 shows statistical results of average gaze moving velocity. Average and standard deviation of this index of each viewer on different stimuli are computed. The large standard deviation on all stimuli indicates that the average gaze moving velocity varies in a wide range over all viewers. Different people view the same stimulus in different ways. As a result, it may be improper to use a metric related to this index as a universal measurement of visual plausibility. One-way repeated measures ANOVA is performed on this index between stimuli simulated with *M1* and corresponding approximating simulation methods in each scenario. No effect is found according to the result. In other words, although people watch different stimuli in different ways, there is no significant difference for individuals based on the type of simulation. Correlation analysis also showed no relation between the gaze velocity and the subjective feeling of quality.

Scenario ID	F-value	Probability
Scenario 1	$F_{1,29} = 7.486103$	$p < 0.015$
Scenario 2	$F_{1,29} = 0.080423$	$p < 0.78$
Scenario 3	$F_{1,29} = 1.438970$	$p < 0.25$
Scenario 4	$F_{1,29} = 2.044929$	$p < 0.20$
Scenario 5	$F_{1,29} = 0.288657$	$p < 0.60$
Scenario 6	$F_{1,29} = 2.333121$	$p < 0.15$
Scenario 7	$F_{1,29} = 0.149406$	$p < 0.75$

Table 5.6: One-way repeated measures ANOVA F-value of average gaze moving velocity. The factor is different simulation methods. ANOVA is performed for each scenario separately. *M5* is not taken into consideration in this result.

Analysis of other eye tracking metrics shows similar results. Thus, we find eye-tracking data helpless in developing an objective metric for evaluating visual plausibility in these simulation scenarios.

5.3.2 Experiment IV.b : The Effect of Complexity of Scenes and Other Factors

We discuss here what factors may affect the authenticity of previously mentioned hypotheses, and what these effects are. We only focus on *Hypothesis I* here. One obvious factor is the number of objects under simulation. *Hypothesis I* is based on a large amount of objects. Intuitively, it may not hold when the number of collisions happening simultaneously is low. This has been witnessed in scenario 2 where the number of simultaneous collisions is lower than other scenarios in the same group. *Hypothesis I* is held true but with lower significance. We thus design a scenario in this phase that measures how the number of simultaneous collisions affects authenticity of *Hypothesis I*. Another factor we will discuss is the sight distance which is the distance from the center of the simulated scene to camera position.

5.3.2.1 The Design of Scenario

We choose to design scenarios based on scenario 4, which has only homogeneous collisions. We thus create a series of similar scenarios in which bunnies are dropped from a certain height and collide with the ground eventually. The number of bunnies dropped at the same time is 1, 4, 16, 64 and 256 in these scenarios. To compare the effect of different simulation methods, *M1*, *M2* and *M5* are used to simulate each scenario. Another scenario with 1000 bunnies simulated is used also. To view all collisions at once, the distance from the simulated scene to the camera in the 1000-bunny scenario is much larger than others, so that it may be difficult for people to discern all the details of each bunny model. This also helps us understand how sight distance can affect *Hypothesis I*.

5.3.2.2 Results and Analysis

The same group of volunteers also participate in this experiment. Viewer rating data is shown in Figure 5.7. One obvious thing we can tell from Figure 5.7 is that number of collisions simulated simultaneously does affect truthfulness of *Hypothesis I*. The average rating for stimuli simulated with *M5* in scenarios with 1, 4, 16 and even 64 objects are uniformly lower than stimuli simulated with the other two methods. As the number of objects increase, the rating for stimuli simulated with this method is closer to that for other stimuli due to more homogeneous collisions. Similar results are seen for *M2*, though

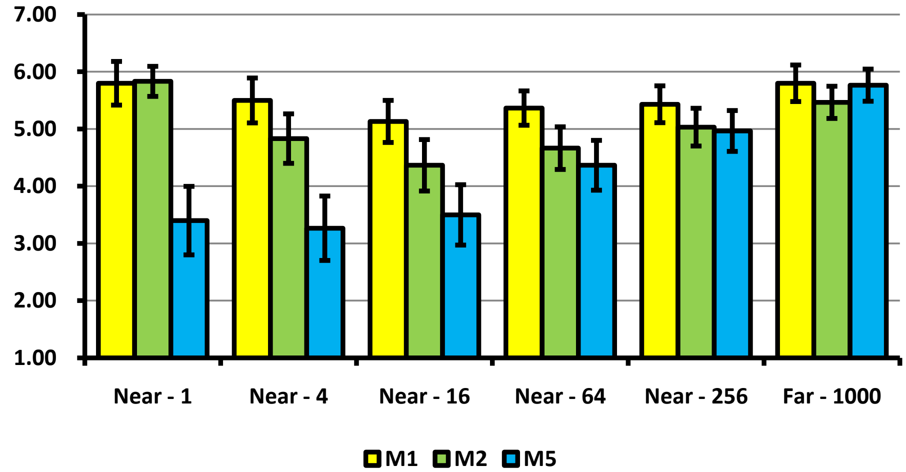


Figure 5.7: Estimation of average and 90% confidence interval of viewers' rating for each stimulus simulated in phase II of experiment. Labels Near - 1 to Near - 256 correspond to scenarios with 1 to 256 bunnies simulated. The sight distance in them is the common value used in all other scenarios. Far - 1000 is a large sight distance scenario.

not as much as for *M5*. The average rating for stimuli simulated with *M2* is uniformly close to stimuli simulated with *M1*. Thus, a conclusion we can draw is that the number of simultaneous collisions does indeed affect the truthfulness of *Hypothesis I*, although this effect may also depend on the simulation method used.

An interesting observation from Figure 5.7 is that when the number of objects is near 16, ratings for stimuli seem to be near minimal for all methods, including *M1*. One possible explanation is that people may be more alert and are more prone to give negative feedback when the number of objects are around 16. This would imply that viewers' ability to compare, track, and discover distortion is best at around 16 objects simultaneously being simulated. We find this to be an interesting initial observation, but further work would be needed before making a rigorous claim.

Finally, ratings for stimuli simulated with all methods in the far sight-distance scenario, which also contains far more objects, are higher than most other scenarios. Because it's more difficult to tell the detail of a simulated bunny model, viewers may accept the assumption that these models are close to particles or spheres. As a result, it's more difficult for viewers to notice computational error. What's more interesting is that in this case the stimulus simulated with *M5* has almost the same average rating as the stimulus simulated with *M1*($F_{1,29} = 0.038718291, p =< 0.85$). To draw any conclusion from this we would need to separate how much of this effect is due to distance of view, and how much is due to number of objects. We expect that both of these are significant factors.

5.4 Study IV Conclusion

We verify the truthfulness of two hypotheses with psychophysical experiments. To our knowledge, our work is the first to verify that the intuition about these hypotheses holds. Furthermore, we also identify four factors which may affect the truthfulness of the hypotheses, and thus need to be taken into consideration when adopting approximated simulation algorithms. Our experimental results also show that the extent of distortion in simulation does not significantly affect viewers' global gaze statistics, as measured with an eye tracker. Thus, eye-tracking does not seem to be a promising avenue for evaluating visual plausibility.

A worthwhile direction for future work is to measure the affecting factors quantitatively both individually and together, to find tolerance thresholds for each. This could help with specific design of simulations, where the simulation method is selected to meet specific criteria. Such a study could help to separate out the contributions of various effects (such as camera distance and number of objects). In addition to those studies, there were also some interesting observations in our study (such as a seeming local minimum when around 16 objects are simulated) that would be worth further study to see how universal the effect is.

6. CONCLUSION

6.1 Summary

We performed four independent studies to investigate several factors which can influence the discrepancy between a physically based animation and its real world counterpart. By influencing the discrepancy, these factors can eventually affect perception of these animations. Through the study of these factors, we have answered the five questions proposed in the introduction. And we further proved the thesis statement based on our conclusions in the four studies.

In Study I and Study II, we proved that the appearance of an object used in the rendering stage can significantly influence the perceived stiffness and the perceived difference of deformation of objects in an animation. We further showed that the influence is dominated by the low-level visual details in the appearance. We proved this by performing three carefully designed experiments. The significant influence claimed is in the statistical sense, which means a confidence level is given for each significant effect. With two other experiments, we tried to measure how the low-level visual details in appearance take effect numerically. We found that the low-level visual details in the form of checkerboard texture with certain spatial frequency and contrast can significantly reduce the perceived stiffness. This discovery gives us more insight as to how the appearance influences the perception of deformation. In both studies, we discovered that increasing the complexity of the scene can reduce the effect of the main factor. In Study I, increasing the number of objects being simulated makes people feel the objects are more rigid and the difference in the perceived stiffness of objects with different appearance is reduced as well. In study II, using a complex background can reduce the effect of the checker board texture on the foreground object.

Given the above observations, it is obvious that simply designing the appearance of an object as it is in real life may not be an optimal choice in controlling the perceived deformability. In simple scenarios, one can reduce or enhance the perceived stiffness of an object by including more low-level visual details or focusing more on the realistic design. One can further tune the effect of low-level visual details by referring to the result of our study and adjusting the spatial frequency and contrast of the visual features in the appearance accordingly. In complex scenarios, to achieve the same perceived stiffness, it is necessary to adjust the low-level visual details accordingly. Because high-level conceptual information becomes more important in complex scenarios, it is reasonable to pay more attention on the design of the high-level information of the appearance.

In Study III and Study IV, we investigates the effect of the resolution of the simulation meshes used in deformable object simulation and the dynamics models used in large scale rigid body simulation. The computational cost of the simulation stage can be controlled by adjusting these factors. It is interesting to people how the perception of an animation can be influenced by adjusting these factors. We proved that the discrepancy between an animation using a simulation mesh with a reduced resolution and an animation using a high resolution mesh may not be noticed by viewer if the reduction is within a certain range. How much the resolution can be reduced without being noticed, however, is influenced by the complexity of the deformation being simulated. Similarly, we proved that the discrepancy caused by approximated rigid body dynamics models in large scale rigid body simulation or the simulation of piling of rigid bodies may not influence the visual plausibility of an animation.

Although the result of our analysis is consistent with some common assumptions, we provide numerical proof to these hypotheses. As an elementary study, we further gave more details of when these hypotheses may not be true. For example, we show that when there is a great deal of curvature in the deformation in the simulation, a reduction of the resolution of the simulation mesh is more likely to be discovered compared to the simulation of other scenarios. Our analysis of these two factors provides more details that people can refer to rather than relying on vague assumptions.

In a general sense, we also showed that the subjective perception of a physically based animation can be influenced by many different factors. Many of these factors can be controlled during the design or the production process. With more knowledge about how these factors take effect, we can better understand how to improve the subjective satisfaction of the animation or how to improve the computational efficiency without harming the subjective satisfaction.

Through the four studies concluded above, we proved the thesis statement:

The factors in the simulation and rendering stages of the animation production pipeline can significantly influence the perception of physically based animation of deformable and rigid objects. These factors can be controlled to enhance or reduce such influence to benefit the animation production.

6.2 Limitations

Because there is very limited previous research on the perception of physically based animation, our studies are all primary so far. The design and the results of our studies are both limited.

First, we make simplifications in our studies to be able to discover statistically significant patterns if there are any. In Study I and II, we mainly use a sphere model and a simple scenario of a free falling single object. Similarly, we only use stimuli with one constrained object being pulled in Study III. In most studies, we only use a single dynamics model. Even in Study IV, we consider only a few models. All these simplifications help us to eliminate influence from other potential factors. However, they also limit the generalization of our observations and discoveries.

Second, some approximations we use in our study could also limit the generalization and soundness of our conclusion. In Study II, we use the checkerboard texture as an approximation of the exponential texture. Although both textures can be used as bases in frequency domain analysis, the energy spectrum of the two textures are different. Thus our measurement on stimuli using a checkerboard texture can not be easily generalized to the exponential texture.

Third, the way we quantify the different aspects of perception is limited. In our studies, we only quantify the perceived stiffness of deformable objects and the visual plausibility of large scale rigid body simulation. Also, we numerically measure the observed difference of deformation. The difference of deformation includes both the perceived stiffness and the perceived DOF. However, our observations based on these quantities are still limited in that they can mainly be used as qualitative references.

Overall, our discoveries are still limited in their applications. Beyond the limitations mentioned above that constrain the generalization of our conclusion and the ones that limit how much numerical reasoning we can make based on the observations, we are still limited in the number of factors investigated.

6.3 Future Work

A primary direction for future work would be to design and perform more specific experiments to verify the correctness of our observations in more generalized scenarios. For example, by using different shapes as the only source of different low-level visual details, we can investigate whether different low-level visual details in appearance, no matter what source they are from, can influence the perception and dominate such influence. Another example is that in Study III, we have only considered constrained objects. We are interested to see if our observations still hold when the objects are allowed to move freely in the scenario.

Another direction of our future work is to investigate applications of our observations to the design process. For example, given the perceived stiffness at different spatial frequency-contrast combinations from Study II, the next question is whether we can predict the perceived stiffness of more complex textures. We can decompose a complex texture into a weighted combination of many single frequency textures using Discrete Fourier Analysis or similar transformations. Knowing the perceived stiffness of these single frequency textures, an intuitive idea is to use the weighted combination of them as a prediction of the overall texture. However, there are still remaining problems to solve. First, is a simple weighted combination of perceived softness of all components a good metric for the overall perceived softness? Second, to predict perceived softness of any basic component from spatial frequency and contrast samples, we may need more samples or more accurate measurement. Third, we need to study if there is discrepancy in people's perception of checkerboard and complex exponential textures. Also, clear high-level hints in composed textures may influence the effect of low-level visual cues. These problems all need to be verified in more experiments. Once we have the predicted perceived stiffness, it can be further used to determine in real time which objects in animation people are less sensitive to. Less accurate models can be used for simulating those objects in a level of detail framework.

Finally, we want to study more factors that people are interested in. For example, dynamics model reduction is another way to improve the simulation efficiency. Similar to mesh resolution reduction, it can potentially influence the visual plausibility of an animation originally simulated using a full dynamics model. People want to know how the reduction of the dynamics model can weaken the visual plausibility. It would be interesting to perform an experiment similar to that in Study III to investigate such influence.

In all, we would like to take a step beyond our conclusions in this dissertation by investigating more factors that can influence the perception of physically based animation and proposing practical applications of our discoveries.

REFERENCES

[1] C. Aliaga, C. O'Sullivan, D. Gutierrez, and R. Tamstorf. Sackcloth or silk?: The impact of appearance vs dynamics on the perception of animated cloth. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception*, SAP '15, pages 41–46, New York, NY, USA, 2015. ACM.

[2] F. Argelaguet, D. A. G. Jáuregui, M. Marchal, and A. Lécuyer. Elastic images: Perceiving local elasticity of images through a novel pseudo-haptic deformation effect. *ACM Trans. Appl. Percept.*, 10(3):17:1–17:14, Aug. 2013.

[3] J. Barbič, F. S. Sin, and D. Schroeder. Vega FEM Library. `http://www.jernejbarbic.com/vega`, 2012.

[4] J. Bender, K. Erleben, J. Trinkle, and E. Coumans. Interactive simulation of rigid body dynamics in computer graphics. In *EUROGRAPHICS 2012 State of the Art Reports*. Eurographics Association, 2012.

[5] H. R. Blackwell. *Psychophysical thresholds: experimental studies of methods of measurement*. Engineering research bulletin. Engineering Research Institute, Univ. of Michigan, 1953.

[6] Blender Online Community. Blender - a 3d modelling and rendering package. `http://www.blender.org`.

[7] D. H. Brainard. The psychophysics toolbox. *Spatial Vision*, 10(4):433–436, 1997.

[8] A. Chapiro, O. Diamanti, S. Poulakos, C. O'Sullivan, A. Smolic, and M. Gross. Perceptual evaluation of cardboarding in 3d content visualization. In *Proceedings of the ACM Symposium on Applied Perception*, pages 47–50. ACM, 2014.

[9] S. Chenney and D. A. Forsyth. Sampling plausible solutions to multi-body constraint problems. In *Proceedings of the 27th Annual Conference on Computer Graphics and Interactive Techniques*, SIGGRAPH '00, pages 219–228, New York, NY, USA, 2000. ACM Press/Addison-Wesley Publishing Co.

[10] T. N. Cornsweet. The staircase-method in psychophysics. *The American Journal of Psychology*, 75(3):pp. 485–491, 1962.

[11] E. Coumans. Bullet Physics Library. `http://bulletphysics.org/`, 2013.

[12] S. J. Daly. Engineering observations from spatiovelocity and spatiotemporal visual models, *In Proceedings of SPIE 3299 Human Vision and Electronic Imaging III*, pages 180-191, San Jose, CA, USA,1998. SPIE.

[13] J. Diemer, G. W. Alpers, H. M. Peperkorn, Y. Shiban, and A. Mühlberger. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in Psychology*, 6:26, 2015.

[14] H. Donald and J. C. Keyser. *Foundations of Physically Based Modeling and Animation*. CRC Press, 2017.

[15] C. Ennis, R. McDonnell, and C. O'Sullivan. Seeing is believing: body motion dominates in multisensory conversations. In *ACM SIGGRAPH 2010 papers*, SIGGRAPH '10, pages 91:1–91:9, New York, NY, USA, 2010. ACM.

[16] I. Fründ, N. V. Haenel, and F. A. Wichmann. Inference for psychometric functions in the presence of nonstationary behavior. *Journal of Vision*, 11(6), May 2011.

[17] M. García, J. Dingliana, and C. O'Sullivan. Perceptual evaluation of cartoon physics: accuracy, attention, appeal. In *Proceedings of the 5th Symposium on Applied Perception in Graphics and Visualization*, APGV '08, pages 107–114, New York, NY, USA, 2008. ACM.

[18] M. García, M. A. Otaduy, and C. O'Sullivan. Perceptually validated global-local deformations. *Comput. Animat. Virtual Worlds*, 21:245–254, May 2010.

[19] S. F. F. Gibson and B. Mirtich. A survey of deformable modeling in computer graphics. Technical report, 1997.

[20] E. Guendelman, R. Bridson, and R. Fedkiw. Nonconvex rigid bodies with stacking. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 871–878, New York, NY, USA, 2003. ACM.

[21] D. Han, S.-w. Hsu, A. McNamara, and J. Keyser. Believability in simplifications of large scale physically based simulation. In *Proceedings of the ACM Symposium on Applied Perception*, pages 99–106. ACM, 2013.

[22] D. Han and J. Keyser. Effect of appearance on perception of deformation. In *Proceedings of the 14th ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, pages 37–44. ACM, 2015.

[23] D. Han and J. Keyser. Effect of low-level visual details in perception of deformation. In *Computer Graphics Forum*, volume 35, pages 375–383. Wiley Online Library, 2016.

[24] J. Harrison, R. A. Rensink, and M. van de Panne. Obscuring length changes during animated motion. *ACM Trans. Graph.*, 23(3):569–573, Aug. 2004.

[25] R. T. Held, E. A. Cooper, J. F. O'Brien, and M. S. Banks. Using blur to affect perceived distance and size. *ACM Trans. Graph.*, 29(2):19:1–19:16, Apr. 2010.

[26] L. Hoyet, R. McDonnell, and C. O'Sullivan. Push it real: perceiving causality in virtual interactions. *ACM Trans. Graph.*, 31(4):90:1–90:9, July 2012.

[27] S.-W. Hsu and J. Keyser. Statistical simulation of rigid bodies. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '09, pages 139–148, New York, NY, USA, 2009. ACM.

[28] S.-W. Hsu and J. Keyser. Piles of objects. In *ACM SIGGRAPH Asia 2010 papers*, SIGGRAPH ASIA '10, pages 155:1–155:6, New York, NY, USA, 2010. ACM.

[29] S.-W. Hsu and J. Keyser. Automated constraint placement to maintain pile shape. *ACM Transactions on Graphics (TOG)*, 31(6):150, 2012.

[30] G. Irving, J. Teran, and R. Fedkiw. Invertible finite elements for robust simulation of large deformation. In *Proceedings of the 2004 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 131–140. Eurographics Association, 2004.

[31] A. Jarabo, T. V. Eyck, V. Sundstedt, K. Bala, D. Gutierrez, and C. O'Sullivan. Crowd light: Evaluating the perceived fidelity of illuminated dynamic scenes. *Comp. Graph. Forum*, 31(2pt4):565–574, May 2012.

[32] T. Kawabe and S. Nishida. Seeing jelly: Judging elasticity of a transparent object. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '16, pages 121–128, New York, NY, USA, 2016. ACM.

[33] D. H. Kelly. Motion and vision. ii. stabilized spatio-temporal threshold surface. *J. Opt. Soc. Am.*, 69(10):1340–1349, Oct 1979.

[34] M. Larkin and C. O'Sullivan. Perception of simplification artifacts for animated characters. In *Proceedings of the ACM SIGGRAPH Symposium on Applied Perception in Graphics and Visualization*, APGV '11, pages 93–100, New York, NY, USA, 2011. ACM.

[35] H. Levitt. Transformed Up-Down Methods in Psychoacoustics. *Journal of The Acoustical Society of America*, 49, 1971.

[36] J. Mannos and D. J. Sakrison. The effects of a visual fidelity criterion of the encoding of images. *Information Theory, IEEE Transactions on*, 20(4):525–536, 1974.

[37] R. McDonnell, M. Breidt, and H. H. Bülthoff. Render me real?: investigating the effect of render style on the perception of animated virtual humans. *ACM Trans. Graph.*, 31(4):91:1–91:11, July 2012.

[38] R. McDonnell, S. Dobbyn, S. Collins, and C. O'Sullivan. Perceptual evaluation of lod clothing for virtual humans. In *Proceedings of the 2006 ACM SIGGRAPH/Eurographics symposium on Computer animation*, SCA '06, pages 117–126. Eurographics Association, 2006.

[39] R. McDonnell, M. Larkin, S. Dobbyn, S. Collins, and C. O'Sullivan. Clone attack! perception of crowd variety. *ACM Trans. Graph.*, 27(3):26:1–26:8, Aug. 2008.

[40] R. McDonnell, M. Larkin, B. Hernández, I. Rudomin, and C. O'Sullivan. Eye-catching crowds: saliency based selective variation. In *ACM SIGGRAPH 2009 papers*, SIGGRAPH '09, pages 55:1–55:10, New York, NY, USA, 2009. ACM.

[41] R. McDonnell and C. O'Sullivan. Movements and voices affect perceived sex of virtual conversers. In *Proceedings of the 7th Symposium on Applied Perception in Graphics and Visualization*, APGV '10, pages 125–128, New York, NY, USA, 2010. ACM.

[42] L. Moody, A. Waterworth, J. G. Arthur, A. D. McCarthy, P. J. Harley, and R. H. Smallwood. Beyond the visuals: Tactile augmentation and sensory enhancement in an arthroscopy simulator. *Virtual Real.*, 13(1):59–68, Feb. 2009.

[43] M. Müller and M. Gross. Interactive virtual materials. In *Proceedings of Graphics Interface 2004*, GI '04, pages 239–246, School of Computer Science, University of

Waterloo, Waterloo, Ontario, Canada, 2004. Canadian Human-Computer Communications Society.

[44] A. Nealen, M. MÃijller, R. Keiser, E. Boxerman, and M. Carlson. Physically based deformable models in computer graphics. *Computer Graphics Forum*, 25(4):809–836, 2006.

[45] J. F. Norman, J. T. Todd, and G. A. Orban. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science*, 15(8):565–570, 2004.

[46] C. O'Sullivan, J. Dingliana, T. Giang, and M. K. Kaiser. Evaluating the visual fidelity of physically based animations. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 527–536, New York, NY, USA, 2003. ACM.

[47] E. G. Parker and J. F. O'Brien. Real-time deformation and fracture in a game environment. In *Proceedings of the 2009 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '09, pages 165–175, New York, NY, USA, 2009. ACM.

[48] J. W. Peirce. Psychopy-psychophysics software in python. *Journal of Neuroscience Methods*, 162(1-2):8 – 13, 2007.

[49] T. Pejsa. Understanding perception of simulated elastic deformations, 2012. Course Project Report.

[50] M. Piovarči, D. I. W. Levin, J. Rebello, D. Chen, R. Ďurikovič, H. Pfister, W. Matusik, and P. Didyk. An interaction-aware, perceptual model for non-linear elastic objects. *ACM Trans. Graph.*, 35(4):55:1–55:13, July 2016.

[51] M. Pražák, L. Hoyet, and C. O'Sullivan. Perceptual evaluation of footskate cleanup. In *Proceedings of the 2011 ACM SIGGRAPH/Eurographics Symposium on Computer Animation*, SCA '11, pages 287–294, New York, NY, USA, 2011. ACM.

[52] P. S. A. Reitsma and C. O'Sullivan. Effect of scenario on perceptual sensitivity to errors in animation. *ACM Trans. Appl. Percept.*, 6(3):15:1–15:16, Sept. 2009.

[53] P. S. A. Reitsma and N. S. Pollard. Perceptual metrics for character animation: sensitivity to errors in ballistic motion. In *ACM SIGGRAPH 2003 Papers*, SIGGRAPH '03, pages 537–542, New York, NY, USA, 2003. ACM.

[54] J. G. Robson. Spatial and temporal contrast-sensitivity functions of the visual system. *J. Opt. Soc. Am.*, 56(8):1141–1142, Aug 1966.

[55] H. Si. *Tetgen - A Quality Tetrahedral Mesh Generator and a 3D Delaunay Triangulator*. Weierstrass Institute for Applied Analysis and Stochastics (WIAS), Mohrenstr. 39, 10117 Berlin, Germany, 2013.

[56] M. A. Srinivasan, G. L. Beauregard, and D. L. Brock. The impact of visual information on the haptic perception of stiffness in virtual environments. In *ASME Winter Annual Meeting.*, 1996.

[57] G. Sweet and C. Ware. View direction, surface orientation and texture orientation for perception of surface shape. In *Proceedings of Graphics Interface 2004*, GI '04, pages 97–106, 2004.

[58] C. D. Twigg and D. L. James. Many-worlds browsing for control of multibody dynamics. In *ACM SIGGRAPH 2007 papers*, SIGGRAPH '07, New York, NY, USA, 2007. ACM.

[59] P. Vangorp, C. Richardt, E. A. Cooper, G. Chaurasia, M. S. Banks, and G. Drettakis. Perception of perspective distortions in image-based rendering. *ACM Transactions on Graphics (SIGGRAPH Conference Proceedings)*, 32(4), 2013. To appear.

[60] M. Vicovaro, L. Hoyet, L. Burigana, and C. O'Sullivan. Evaluating the plausibility of edited throwing animations. In *Proceedings of the ACM SIGGRAPH/Eurographics*

Symposium on Computer Animation, SCA '12, pages 175–182, Aire-la-Ville, Switzerland, Switzerland, 2012. Eurographics Association.

[61] E. Vul, G. Alvarez, J. B. Tenenbaum, and M. J. Black. Explaining human multiple object tracking as resource-constrained approximate inference in a dynamic probabilistic model. In *Advances in Neural Information Processing Systems*, pages 1955–1963, 2009.

[62] W. H. Warren, E. E. Kim, and . Husney, R. The way the ball bounces: visual and auditory perception of elasticity and control of the bounce pass. *Perception*, 16(3):309–336, 1987.

[63] H. Winnemöller, D. Feng, B. Gooch, and S. Suzuki. Using npr to evaluate perceptual shape cues in dynamic environments. In *Proceedings of the 5th International Symposium on Non-photorealistic Animation and Rendering*, NPAR '07, pages 85–92, New York, NY, USA, 2007. ACM.

[64] H. Yee, S. Pattanaik, and D. P. Greenberg. Spatiotemporal sensitivity and visual attention for efficient rendering of dynamic environments. *ACM Trans. Graph.*, 20(1):39–65, Jan. 2001.

[65] T. Y. Yeh, G. Reinman, S. J. Patel, and P. Faloutsos. Fool me twice: Exploring and exploiting error tolerance in physics-based animation. *ACM Trans. Graph.*, 29(1):5:1–5:11, Dec. 2009.

[66] L. Zhu, X. Hu, and L. Kavan. Adaptable anatomical models for realistic bone motion reconstruction. In *Computer Graphics Forum*, volume 34, pages 459–471. Wiley Online Library, 2015.